Using Watson Perceptual Model to Improve Quantization Index Modulation Based Watermarking Schemes

Qiao LI

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of the

University of London.

Department of Electronic and Electrical Engineering

University College London

September, 2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration

I, Qiao LI, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Quantization index modulation (QIM) is a popular watermarking scheme that has received considerable attention. Nevertheless, there are practical limitations of QIM. For example, traditional QIM uses a fixed quantization step size, which may lead to poor fidelity in some areas of the content. More serious problems of the original QIM algorithm include its extremely sensitivity to valumetric scaling (e.g., changes in amplitude) and re-quantization (e.g., JPEG compression).

In this thesis, we first propose using Watson's perceptual model to adaptively select the quantization step size based on the calculated perceptual "slack". Experimental results on 1000 images indicate improvements in fidelity as well as improved robustness in high-noise regimes.

Watson's perceptual model is then modified such that the slacks scale linearly with valumetric scaling, thereby providing a QIM algorithm that is theoretically invariant to valumetric scaling. In practice, the robustness against valumetric scaling is significantly improved, but scaling can still result in errors due to cropping and roundoff that are an indirect effect of scaling. Two new algorithms are proposed — the first based on regular QIM and the second based on rational dither modulation. A comparison with other methods demonstrates improved performance over other recently proposed valumetric-invariant QIM algorithms, with only small degradations in fidelity.

Spread transform dither modulation (STDM) is a form of QIM that is more robust to re-quantization. However, the robustness of STDM to JPEG compression is still poor and it remains very sensitive to valumetric scaling. We describe how a perceptual model can be incorporated into the STDM framework to (i) provide robustness to valumetric

scaling, (ii) reduce the embedding-induced perceptual distortion and (iii) significantly improve the robustness to re-quantization.

Acknowledgements

First and foremost I wish to thank my supervisor Prof. Ingemar Cox, director of the UCL Adapstral Park. I remember that on the first day I was at UCL, he gave us a guided tour of London and I clearly remember that on Embankment Bridge where we were looking at London Eye, nicely he said "You have three years to enjoy". Well, it ends up with longer time, but I did enjoy, of course not only London but also work under his supervision. From him, I learned how to think, how to do, and how to present a research. I have benefited from that and will continue to benefit for a long time. I can not give enough praise that Prof. Cox deserves because it would take the whole thesis! I can simply say, without his support, I would not be at this point.

I would like to gratefully thank my secondary supervisor, Dr. Gwenaël Doërr. There are countless times when his suggestions help me out of trouble. He is always supportive and his talk is always informative. I would also thank some external experts: J. C. Oostveen, T. Kalker and M. Staring who provided the code to their algorithm, F. Perez-Gonzalez who gave insights into their method.

I wish to express my gratitude to all faculty and staff at UCL Adastral Park, especially Prof. Fred Stentiford, Dr. Shi Zhou, and Dr. Kai-Kit Wong who have provided inspiring discussions, and Neil Marjoram and Rich Hutchinson who I have asked for help on IT systems. I also owe many thanks to my colleagues and friends at UCL: Chin Wang, Lin Lin, Vishwa Vinay, Ade Bamidele, Shijie Zhang. Writing this has reminded me many people and stories although it is impossible to make a full list here.

Finally, the greatest thanks go to my parents. It is their endless love that makes me overcome difficulties whenever I meet in my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A couple of decades ago most audio, image and video were traditionally stored and transmitted in analog formats. For example, prints, tapes and cinema films were frequently used. Analog signals in all these forms may have noticeable degradations when they are copied, transmitted or stored for a long period. The rapid growth of digital information technology means that many of these analog formats have been or are being replaced by digital formats.

Digital media offer several distinct advantages over analog media: the quality of digital audio, image, and video signals can be higher than that of their analog counterparts. Editing is easy because one can access the exact discrete locations that should be changed. Copying is simple and result in no loss of fidelity, i.e., a copy of a digital media is identical to the original. And digital audio, image, and videos are easily transmitted over networked information systems.

Meanwhile, the emergence of high bandwidth networks has facilitated Peer-to-Peer (P2P) file-sharing technology and other systems to transfer audio, image, video files without copyright permission.

The protection and enforcement of intellectual property rights for digital media has therefore become an important issue. Concerns related to digital piracy have primarily motivated research for digital watermarking, which is basically a technique to embed a message, for instance, copyright information, into a digital copy of a song, video or picture.

In the last ten years, digital watermarking has attracted attention because of its applications to, for example, broadcast monitoring, owner identification, transaction tracking, content authentication, and copy control. All these applications have motivated research in digital watermarking.

1.1 Basics of Watermarking

The digital watermarking techniques presented in this thesis concentrates on the watermarking of electronic signals. A specific song, video, or picture - or a specific copy of such - is referred to as a *Work*. And a set of all possible Works is referred to as *content*. The central idea of digital watermarking is to embed information about a Work into the original Work itself to generate a watermarked Work. The embedding is desired to be done in such a way that the original Work (also referred to as the host signal) and the watermarked Work (watermarked signal) have no perceptual difference. In general, watermarking systems consist of two processes: embedding and detection. A simple yet complete model of a watermarking framework appears in Figure 1.1.

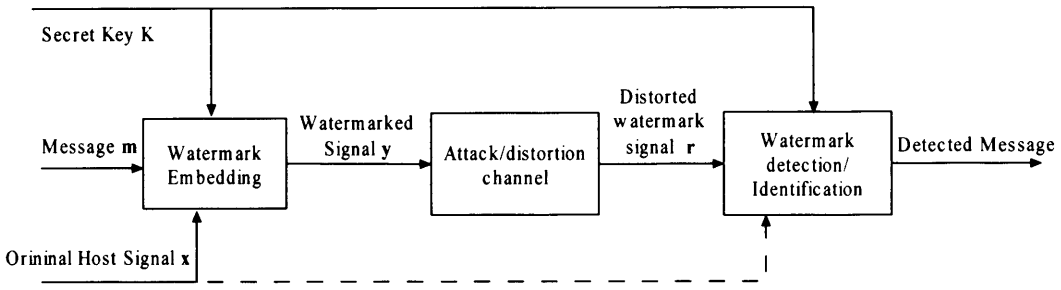


Figure 1.1: Framework of digital watermarking

Here, the solid lines are the necessary elements and operations while the dashed line is optional. The embedding process takes the host signal, a secret key and the message to be embedded and then produces the watermarked signal. The decoding process takes the received distorted watermarked signal and a secret key, and attempts to detect if the image contains a watermark, and if it does, decodes the message. Note that this diagram shows that the host signal may or may not be available to the detector, which determines if the detector is an informed detector or a blind detector. This issue

is further discussed in Section 2.2.4.

A host signal can be an image, 3D graphic, video, music, text document or html document. A survey of methods used to hide information in various media is given in [PAK99a]. In this thesis, we mainly focus on images. However the framework can be generalized to other media.

1.2 Organization of the Thesis

This thesis is organized as follows: Chapter 2 firstly gives a general background of digital watermarking: applications and properties. It then briefly reviews digital watermarking schemes from a communication's perspective, particularly, communication with side information. Here, some popular watermarking algorithms, such as spread spectrum watermarking, dirty paper trellis watermarking and lattice based watermarking, are briefly introduced. In later part of Chapter 2, we move on to perceptual models. Perceptual models are very important to digital watermarking, not only because they can measure the quality and differences between of multimedia signals, but also because they can help improve watermarking designs. This chapter begins with the importance of perceptual models and then discusses several visual models. In particular, we study the Watson's perceptual model [Wat93a]. Its main features are described both qualitively and quantitatively for later use.

As a lattice code based watermarking scheme, quantization index modulation (QIM) is popular because of its ease of implementation, computational efficiency and amenability to theoretical analysis. Chapter 3 describes the principles of QIM and its extensions such as dither modulation, distortion-compensated QIM and spread transform dither modulation. Traditional QIM also has some practical limitations. For example, it may introduce perceptually visible distortion in some areas of the content. Even worse, QIM is extremely sensitive to valumetric scaling (e.g., changes in amplitude) and re-quantization (e.g., JPEG compression). These problems are discussed and illustrated by experimental results in the last section of this chapter.

Chapter 4 attempts to solve the problem of QIM's vulnerability to valumetric scal-

ing and re-quantization. We first propose using Watson's perceptual model to adaptively select the quantization step size based on the calculated perceptual "slack". It indicates improvements in fidelity as well as improved robustness in high noise regimes of additive noise. Second, Watson's perceptual model is modified such that the slacks scale linearly with valumetric scaling, thereby providing a QIM algorithm that is theoretically invariant to valumetric scaling. In practice, the robustness against amplitude scaling is significant, but scaling can still result in errors due to cropping and roundoff that are an indirect effect of scaling. Rational dither modulation (RDM) [PGBAM04] has been proposed to provide valumetric invariance to QIM. Chapter 4 also illustrates how we combine RDM with our modified Watson's model to obtain further improvements. Spread transform dither modulation (STDM) is a form of QIM that is more robust to re-quantization. However, the robustness of STDM to JPEG compression is still very poor and it remains very sensitive to amplitude scaling. In the last part of Chapter 4, we show how a perceptual model that scales linearly with amplitude scaling can be used to (i) provide robustness to amplitude scaling, (ii) reduce the perceptual distortion at the embedder and (iii) significantly improve the robustness to re-quantization.

Experimental results in Chapter 5 evaluate performance of our methods and confirm the improvements. The experiments include tests on both 1,000 images from Corel database and some standard images (for example Lena, Barbara Cameraman). Sample images and a user trial are also illustrated in Chapter 5

Finally, Chapter 6 summarizes our work. In addition, some directions for future work are suggested.

Chapter 2

Overview of Digital Watermarking and Perceptual Models

This chapter begins with some applications of digital watermarking. Necessary properties of watermarking are then described. Next, a framework that models watermarking as a communications problem is discussed and terms and concepts used in this thesis are described. This is followed by a review of digital watermarking algorithms and developments. later in the chapter, some perceptual models are discussed.

2.1 Applications of Digital Watermarking

Because digital watermarking is imperceptible, and survives common format conversions and signal processing, it has be used for a variety of applications, described next.

Initially, digital watermarking technology attracted significant research attention because it can be used for copyright related issues of digital contents. Presently, watermarking has also been used for application beyond copyright related issues. Generally speaking, when an additional information is related to a Work, it can be embedded into the Work as a watermark.

2.1.1 Transaction tracking

In transaction tracking, the owner of a Work embeds a unique watermark into each copy provided to a consumer. If illegal copies of the Work are subsequently found, they can then be analysed using a watermark detector. Detection of the watermark identifies the

source of the illegal copy.

For example, there are several cases in recent years where Hollywood studios have used digital watermarks to identify illegal copies of *Oscar* screeners [Bar, the, Bet]. A movie screener normally refers to a copy which is sent, sometimes in advance of general release to public, to the awards voters, who are members of the *Academy of Motion Picture Arts and Sciences*. There are $5,830$ voting members as of 2007. Unfortunately, illegal copies of these movie screeners may be distributed. In recent years, transaction tracking watermarks were embedded in each of the screeners. These watermarks have successfully identified the sources of illegal copies.

2.1.2 Owner identification

Traditionally, textual copyright notices, such as "© *data owner*", are used to identify the copyright of a Work. However, they can be easily removed from a Work. For example, the portion of an image which includes the copyright notice might be cropped off. Another problem with textual notices is that they are visible and may be annoying if they cover a portion of an image. While it is possible to put them in less important areas of an image, this makes visible copyright notices more susceptible to being cropped. The situation for audio Work is even worse since the copyright notice is marked on the packaging or physical medium such as the disk, cassette, etc. These notices may no longer exist when the audio content is copied to another physical medium. In fact, for audio content only in electronic form, e.g. on a website, there is no physical medium or packaging.

Digital watermarks are both imperceptible and not easily removed from a Work. They can therefore offer complimentary performance to text for owner identification. Using a watermark detector, users of Works can obtain the watermark message to identify the owner, even after the Work has been modified in ways in which the textual copyright notices would be removed.

2.1.3 Broadcast monitoring

Commercial advertisements broadcast on TV are very expensive. Advertisers who purchase airtime want to ensure that they receive the time they pay for from the broadcaster. Owners of contents are also keen to know if their Work is illegally re-broadcast by other stations.

Digital watermarking is one way to implement *active broadcast monitoring*. Using watermarking, identification information can be embedded within the content itself. Given a received signal, a corresponding detector in the system first detects the watermark and then records how often and how long the advertisement was broadcast.

2.1.4 Copy control

The applications of watermarking we discussed so far deter improper use or can be used to investigate improper use. For instance, transaction tracking helps to find who is the source of illegal copies, but is does not prevent illegal copies being made. For those applications, watermarking technique serves as a deterrent against wrongdoing.

Obviously, it is better to prevent illegal actions. Encryption is a technique to prevent illegal copying. Once encrypted, the Work is unusable to anyone who does not have a valid key. However, a central weakness of encryption is that the content must be ultimately decrypted to be used, and once decrypted, there is no more protection for the content. An adversary need only sign up as a customer to obtain a valid key, and then can easily pirate the content after it is decrypted.

Digital watermarking provides a complementary method of copy control, since digital watermarks embedded into the content are always present together with the content [BCK+99]. For example, a recording device equipped with a watermark detector could prohibit recording whenever a *never-copy* signal as a watermark is detected. This feature is referred to as *record control*. More discussions of this can be found in [CMB01].

2.1.5 Content authentication

It is very easy to tamper with digital content. For example, a photo can be easily modified by software, such as Photoshop. This may result in alterations that are difficult to be perceived or detected. However, in some circumstances, such as a news report, it can be problematic if the interpretation of the photo is affected (e.g. removal of a person from a photo). Thus there is a need for authentication of the integrity of the content.

A digital signature is a cryptographic approach for the purpose of content authentication. It is created through the use of a hash function. Signatures can be embedded into content as digital watermarks. They are referred to as *authentication marks* which include *fragile watermarks* and *semi-fragile watermarks*. A fragile watermark, for example a least-significant-bit (LSB) based watermark, is designed to verify exact integrity, i.e. the watermark becomes invalid after the watermarked Work is modified in any way. On the other hand, a semi-fragile watermarking is required to survive legitimate distortions, such as JPEG compression, but is invalidated by illegitimate manipulations. Readers are directed to [HZG04, SC05, CMB01, LPD00, LC00, WD99, WD96, CMT+99, WL98, WK00, FA00, BS99, LT98] for more information on watermarking authentication.

2.1.6 Legacy enhancement

For a widely used and largely deployed system, upgrading to provide new features or improved functionalities can be difficult. Ideally, an upgrade should be done in such a way that the new system is backward compatible, i.e., it continues to work with the existing legacy system. Digital watermarking is one of the ways in which to maintain compatibility.

An example of the use of watermark to upgrade an existing system is the Tektronix's synchronization technology. Digital signal processing may introduce different delays to audio and video signals. This can lead to a disconcerting phenomenon often referred to as lip-sync. In this case, it is clearly noticed that the motion of the lips is

either ahead or behind the speech. Tektronix has developed a watermark encoder to synchronize audio and video signals. This product firstly embeds a compressed version of an audio signal into the video signal, prior to any signal processing. Once all signal processing is done, the processed audio signal is compared with the audio signal extracted from the video. The comparison helps to determined the time delay between audio and video and then the delay can be corrected.

There is another similar application: when an mp3 player attempts to display the lyrics in synchrony with the songs. Lyrics can be embedded into audio signals for synchronizing purpose. This technology, pioneered by MarkAny of Korea, is known as MediaSync.

Other proposed system enhancements include air traffic control. In order to upgrade the system and maintain compatibility, Eurocontrol, the European Organization for the Safety of Air Navigation, has considered inserting watermarks into pilots' voice communications between aircrafts and ground-based control sites [HHK03, HHK04, HH05b,HH05a]. These watermarks provide a digital identifier, e.g., an aircraft's unique tail number. The proposed system is compatible with the existing equipment in both aircrafts and ground-based air traffic control centers so that it can be gradually introduced.

2.2 Properties for Digital Watermarking

Each watermarking application has its own specific requirements. A number of defining properties for watermarking systems can be characterized [CM97, HK99, PAK99b, WPD99], some of which are highlighted in this section.

2.2.1 Robustness

Robustness refers to the ability of the inserted watermark message to withstand common signal processing or modifications such as contrast/brightness adjustment, noise addition, lossy compression, quantization, D-A/A-D conversion, etc. In most cases, a watermark needs only survive those signal operations likely to occur after embedding and before detection. Clearly, this is application dependent. Thus requirements of wa-

termarking robustness vary according to certain applications. In the case of television broadcast monitoring, the watermark needs to survive the transmission process which may involve lossy compression, D-A conversion, low-pass filtering, and additive noise. On the other hand, a watermark for this application need not survive, for example, rotation, scaling.

In some cases, robustness of a watermark is *undesirable*. For example, as discussed in 2.1.5, *fragile watermarking* is used for authentication purposes so that any signal processing should result in losing the watermark. Fragile watermarking is discussed in [WD99, CMT+99], but is outside the scope of this thesis. Instead, we mainly focus on how to design robust watermarking techniques. For these robust schemes, a watermark message needs to be accurately extracted from a distorted Work.

2.2.2 Fidelity

In digital watermarking, it is often required that a watermarked Work be indistinguishable from its host signal (original Work). An embedded watermark message should not interfere perceptually with the host signal. We use the term *fidelity* to designate this concept as the perceptual similarity between the watermarked Work and the original Work. In some publications the word *transparency* is also used. Another important term is *embedding-induced distortion* which represents the distortion introduced by embedding a watermark message into an original Work. (i.e. the "difference" between the original and the watermarked signal). To maintain fidelity, the embedding-induced distortion is required to be as low as possible. The property of fidelity is often seen conflicting with robustness. For example, to achieve higher robustness, we can increase the watermark strength which makes more modifications to the host signal. However, larger modifications will introduce more distortions that may be perceptible and fidelity is therefore reduced.

It is reasonable that higher fidelity watermarks are desired for high quality applications such as HDTV or DVD. For relatively low quality broadcast technologies such as NTSC TV or AM radio, lower fidelity is acceptable since it is then harder to per-

ceive the differences between the original and the watermarked Works. Sometimes, even mildly perceptible watermarks are acceptable in exchange for higher robustness, capacity or lower cost. For example, Hollywood dailies are raw materials only shown to those involved in film production. Consequently, a small visible distortion by a watermark will not diminish their value.

2.2.3 Capacity

Capacity refers to the amount of message bits to be embedded into a host signal. For images, capacity is defined according to the number of bits embedded per pixel or a fixed size of an image. For audio, capacity refers to the number of embedded bits per second. For video, capacity refers to the number of bits to be embedded either per frame or per second. There is also a trade-off between the two properties of robustness and capacity. For example, one can increase the capacity by decreasing the number of samples allocated to each bit to be embedded, but this is counterbalanced by a loss of robustness.

The required watermarking capacity varies for different applications. Copy control applications may only need 4 bits of information to be embedded over a period of, say, every 10 seconds for music or 1 minute for a video segment. Under these circumstances, the capacity is about 0.5 bits per second for music and 0.07 bits per second for video, respectively. By contrast, much larger number of bits may be required every second for television broadcast monitoring to identify all commercials [CMB01].

2.2.4 Blind or informed detection

A watermark detector which requires access to the original, unwatermarked Work is referred to as an *informed detector*. Detectors that do not need any information of the original Work are called *blind detectors*. Whether a watermarking system employs blind or informed detection is a critical factor to determine which application it can be used for.

Informed detection is only practical in those applications where the original Work is available. Consider the case of transaction tracking. It is most likely to be the owner

of the original Work who runs the detector in an effort to discover who leaked the source of an illegal copy. The owner can provide the original Work to the detector to extract the watermark from the illegal copy. By doing so, the detection performance may be considerably improved. For example, when rotation or amplitude scaling is applied to a watermarked signal, informed detection allows angle and scaling factors to be estimated and removed. However, in many applications, for instance copy control, detection must be performed without access to the original Work. This thesis mainly discusses blind detection systems.

2.2.5 Security

When a watermark is used to provide enhanced features to consumers, there may not be an adversary. These watermarks need not to be secure against any type of attack. However, many applications do require some level of security. The *security* of a watermark is its ability to withstand hostile attacks. A hostile attack is any process specifically intended to remove or alter the watermark. Most types of attacks can be broken into three categories: *unauthorized removal, unauthorized embedding and unauthorized detection.* Unauthorized removal and embedding are also referred to as *active attacks* since they modify the watermarked Work. In contrast, unauthorized detection does not modify the Work therefore it is referred to as a *passive attack.*

Unauthorized removal describes attacks that prevent watermark from being detected. Note that this attack does not necessarily mean reconstructing the original Work. Rather, an adversary may generate a new Work which is perceptually similar to the original Work but from which a watermark cannot be detected. A *collusion attack* [TWWL03, EKK99, KLM+97, Sto96] is one form of unauthorized removal. In a collusion attack, the adversary first obtains several copies of a given Work, each containing a different watermark, and then combines them to produce a copy with no watermark. This is a big concern in an application such as transaction tracking, which embeds a different watermark into each copy.

Unauthorized embedding, also known as *forgery*, refers to embedding illegitimate

watermarks into Works that should not contain them. Consider the case of owner identification, if an adversary has the ability to perform unauthorized embedding, the detector may falsely identify the owner of the Work.

There are also applications in which the unauthorized detection, a passive attack, is an important issue. For example, suppose that there is a broadcast monitoring company which embeds watermarks for free but charges for the monitoring reports. An adversary who can perform unauthorized detection could set up a rival company selling monitoring reports without incurring the cost of watermark embedding.

2.2.6 Cost

The expense of deploying watermarking systems can be a very complicated topic and depends on the business models involved [Dec01]. The issues of concern include speed to perform embedding and detection, the number of embedders and detectors that must be deployed, and the complexity of hardware and software implementations.

Regarding the cost of computing speed, in the application of broadcast monitoring, the watermarking systems must keep up with the real-time broadcast, otherwise it fails to monitor the program continuously broadcast. By contrast, a detector used to identify ownership is still valuable even if it needs a couple of days to extract a watermark message.

2.3 Model of Digital Watermarking

This thesis so far has discussed watermarking at a fairly non-technical level. Here, we begin to explore technical principles in more detail. To help address these issues, this section defines some notations necessary to begin our technical discourse, and that are used throughout this thesis.

Scalars Scalar values and individual members of sets are denoted in italic lower case: x, y and so on. The magnitude of scalar value x is denoted as $|x|$.

Sets Sets are denoted in a calligraphic font. For example, the set of real numbers is $\mathcal{R}$, a set of messages is $\mathcal{M}$ and the number of elements in set $\mathcal{M}$ is denoted $|\mathcal{M}|$.

Vectors and arrays One-dimensional vectors and higher-dimensional arrays appear in this thesis as boldface, lowercase letters; as, for example, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{m}$. An sample of vector $\mathbf{x}$ is marked as x_i, where the subscript i is the index of x_i in the vector. In addition, the Euclidean length or the magnitude of a vector $\mathbf{x}$ is denoted as $|\mathbf{x}|$. The *mean* of a vector denoted $\bar{\mathbf{x}}$, is the average of its elements. The *variance* of vector $\mathbf{x}$ is written as σ_x^2.

A simple watermarking system can be modeled as illustrated in Figure 2.1.

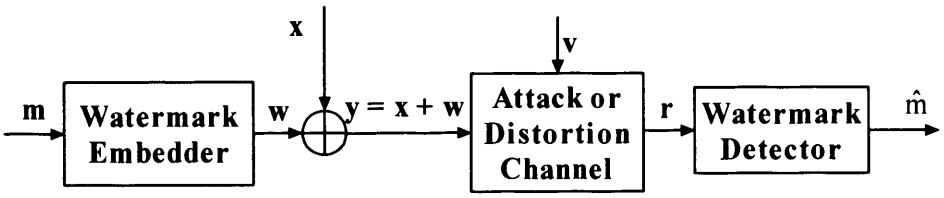


Figure 2.1: A model of digital watermarking

Here, the message, $\mathbf{m}$, is input into the watermark embedder, which outputs a watermark, $\mathbf{w}$ that is added to the original Work or host signal, $\mathbf{x}$, (i.e. image or song), to produce the watermarked Work, $\mathbf{y}$. The watermarked Work then undergoes a number of distortions that are modeled as an unknown noise source, $\mathbf{v}$. The watermark detector receives a distorted, watermarked Work, $\mathbf{r}$, i.e. $\mathbf{r} = \mathbf{x} + \mathbf{w} + \mathbf{v}$ and decodes a message $\hat{\mathbf{m}}$.

We now define some terms. The *embedding rate* is the number of watermarking message bits embedded in each host signal sample. For example, consider embedding a watermark message into an image, if there are 4096 bits embedded into a 512×512 image, one bit is embedded into 64 pixels on average. The embedding rate is therefore one bit per 64 pixels. Importantly, note that this does not necessarily mean 64 pixels are modified to carry a single bit.

Mean square error (MSE) is usually used to measure the embedding-induced distortion between the host signal $\mathbf{x}$ and the watermarked signal $\mathbf{y}$. It is defined as:

$$MSE = \frac{1}{L}\sum_{i=1}^{L}(y_i - x_i)^2 \tag{2.1}$$

where L is the length of the signal vectors $\mathbf{x}$ and $\mathbf{y}$, that is, the total number of signal samples to be measured. Here, x_i and y_i denote a sample of the original signal and the watermarked signal, respectively.

If we define the difference between the watermarked signal $\mathbf{y}$ and the original signal $\mathbf{x}$ as the watermark, $\mathbf{w} = \mathbf{y} - \mathbf{x}$, the document-to-watermark ratio (DWR) is defined by:

$$DWR = 10 \log_{10} \left(\frac{\sigma_x^2}{\sigma_w^2} \right) \tag{2.2}$$

where σ_x^2 and σ_w^2 are the variances of $\mathbf{x}$ and $\mathbf{w}$, respectively.

For watermarking system, the phrase peak signal-to-noise ratio, often abbreviated to PSNR, is defined as the ratio between the maximum possible power of a host signal and the power of watermark.

$$PSNR = 10 \log_{10} \frac{MAX_x^2}{\frac{1}{L} \sum_{i=1}^{L} (w_i)^2}, \text{ where } w_i = y_i - x_i \tag{2.3}$$

Here, MAX_x is the maximum pixel value of the image. When the pixels are represented using 8 bits per sample, this is 255.

Similarly, the watermark-to-noise ratio (WNR) is defined as the ratio of the power of the watermark to the power of the noise which is added to the watermarked signal before watermark detection.

$$WNR = 10 \log_{10} \left(\frac{\sigma_w^2}{\sigma_v^2} \right) \tag{2.4}$$

To measure the performance of watermarking detection, the *bit error rate* (BER) is often used, which is the percentage of bits that have errors relative to the total number of bits embedded as a watermark message.

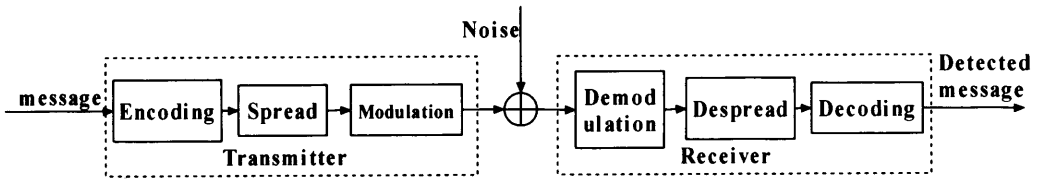
2.4 Spread Spectrum Watermarking

Signal jamming is a deliberate effort by an adversary to disrupt communications between two or more parties. This is a big threat to military communications. *Spread-*

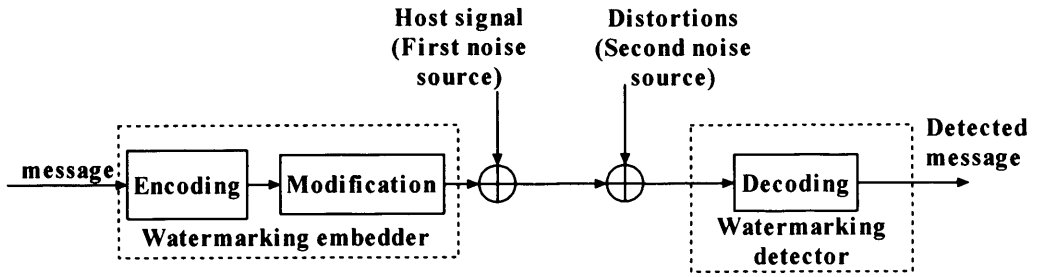
spectrum (SS) communication initially emerged as a method of military communication, which is less sensitive to interference or jamming. In SS communication, the desired signal is spread over a much wider bandwidth than would normally required [PSM82]. One early and simple example of SS technology is known as *frequency hopping* [MA42]. In this system, signals are spread through a large number of frequency bands and each frequency band only carries a fraction of the signal. This spreading is carried out using a pseudo-random sequence. It is called "Pseudo" because the sequence is determined by a secret key and can be re-generated once the key is known. The exact form of this spreading is a secret only known to the signal sender and receiver. Without the key of the spreading function, it is very difficult for an adversary to detect or interfere with the signal.

The problem of watermarking bears a striking similarity to a communication problem and it has been widely recognized that digital watermarking can be studied by communication theories. Figure 2.2 illustrates a spread-spectrum communication system and a digital watermarking system. It can be seen that the watermark embedder plays the role of the channel transmitter, and the watermark detector plays the role of the receiver. The objective of both systems is to transmit a message reliably. Note that in Figure 2.2(b), the watermarking system is considered to have two noise sources. The first one is the host signal (e.g. an image) and the second one is the distortions between embedding and detection.

As discussed in Section 2.2, some common properties of watermark, such as fidelity and robustness, are contradictory to one another. For example, fidelity implies that the watermark signal carried by the Work have low power to avoid perception, while robustness requires that the watermark have large power for better detection. Spread spectrum communications can be used to address this trade-off between fidelity and robustness. First, the signal energy carried by one particular frequency band is small, which reduces the risk of perceptible artifacts. Second, despite the low energy signal in any single frequency band, there may still be high energy at the receiver if the energy in each frequency is summed (or de-spread).



(a) Spread-spectrum communication



(b) Digital watermarking

Figure 2.2: Spread-spectrum communication and digital watermarking

The ideas of spread spectrum communications were combined with watermarking framework to develop spread spectrum watermarking technique [CKLS96, BGML96, SZT96, Ó RuanaidhDB96, PZ98]. Cox *et al.* [CKLS97] used a watermark consisting of a zero-mean unit variance Gaussian of sufficiently low power and embedded it in the low (non-DC) DCT coefficients of images. Thus the watermark is spread over a set of visually important frequency components. This embedding function is represented in Equation (2.5),

$$\mathbf{y} = \mathbf{x} + \alpha \mathbf{w_m} \tag{2.5}$$

where $\mathbf{x}$ is the host signal, $\mathbf{y}$ is the watermarked signal, and $\mathbf{w_m}$ is a pseudo-randomly coded reference vector specific to a watermarking message, $\mathbf{m}$. The factor α can be used to adjust the watermark strength.

Detection of the watermark is performed using a similarity measure, similar to a correlation detector. Given a received signal $\mathbf{r}$, the correlation, z, is computed as shown in Equation (2.6),

$$z = \mathbf{r} \cdot \mathbf{w_m} = (\mathbf{x} + \alpha \mathbf{w_m} + \mathbf{v}) \cdot \mathbf{w_m} \tag{2.6}$$

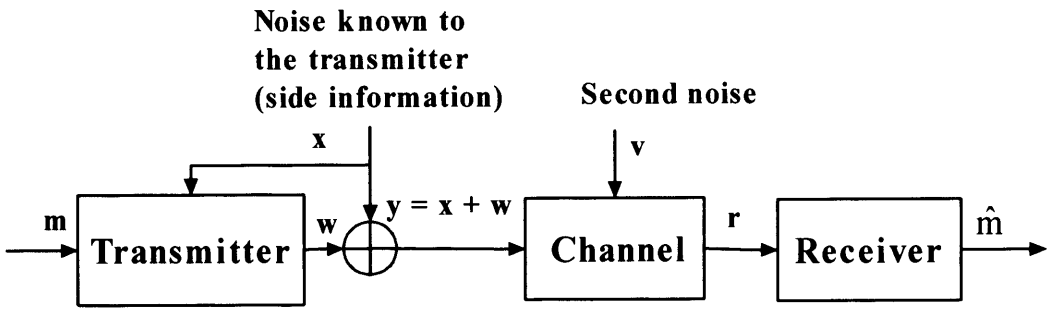
where $\mathbf{v}$ is the noise added after embedding. The detected message is then determined

by judging which code word has the highest correlation to the received signal **r**.

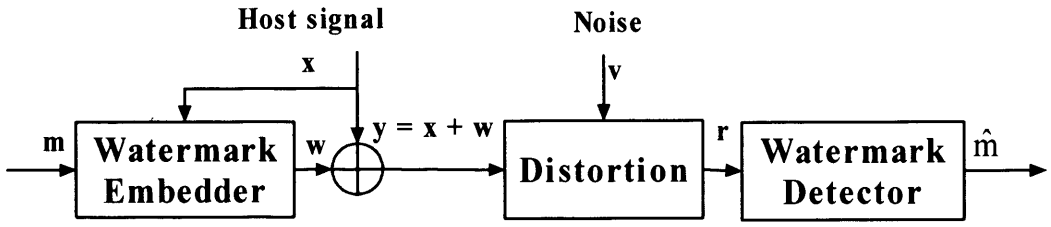
Spread-spectrum watermarking has received considerable attention. Interesting extensions can be found in [PZ98, Ó RuanaidhP98, MF03].

2.5 Watermarking as a Problem of Communication with Side Information

In 1983, Costa published a paper entitled "Writing on Dirty Paper" [Cos83]. It studied the capacity of a Gaussian channel with two noise sources. The first noise is completely known to the transmitter but unknown to the receiver. Meanwhile, neither the transmitter nor the receiver knows the second noise source. The first known noise source can be referred to as *side information* and the system is often called communication with side information. Digital watermarking can be analyzed as a problem of communication with side information, as illustrated in Figure 2.3.



(a) Model for Costa's dirty paper communication



(b) Model for digital watermarking

Figure 2.3: Digital watermarking as a problem of communication with side information

Here in Figure 2.3(b), the host signal, **x**, can be viewed as a noise signal, which

is known to the embedder (transmitter) but not to the detector (receiver). Signal $\mathbf{w}$ is generated to transmit a message $\mathbf{m}$. The Gaussian channel noise is denoted as $\mathbf{v}$, and the received data is $\mathbf{r} = \mathbf{x} + \mathbf{w} + \mathbf{v}$. Costa presented the interesting result that the capacity of this system is the same as if the known noise source, $\mathbf{x}$, was not present. Thus the host interference is eliminated.

In the late 1990's, researchers [CMM99, CPR99, BG99b, MCB00] realized the benefit of modeling watermarking as communication with side information. According to Costa's dirty paper theory, the host signal interference can be eliminated since it is the known noise (side information) to the embedder. This realization has motivated more effective watermarking algorithms based on *informed coding*.

A watermarking embedder needs to generate a code word to represent the given message to be embedded. In *blind coding*, the watermark message is represented with a code word which is independent of the host signal. In other words, the encoding process does not make any use of the available host signal. The word *blind* is used to emphasize this fact. For a given message $\mathbf{m}$, the same code word will always be generated. Thus it is a one-to-one mapping between a message and a codeword. In *informed coding*, by contrast, there is a one-to-many mapping between messages and code words. Each message can be mapped into a set of alternative code words and the choice of which codeword to use is determined by the host signal, i.e., side information. This one-to-many mapping is sometimes referred to as dirty paper coding.

Informed coding for watermarking systems include dirty paper trellis codes and lattice codes, which will be briefly described in Section 2.6.

2.6 Dirty Paper Watermarking Schemes

Costa's original dirty paper codes required exhaustive search on an extremely large number of codewords. That is, Costa's dirty-paper theory did not provide a practical method of implementation. Dirty paper trellis codes [MDC02, MDC04] and lattice codes [BG01a] are two important types of approaches to permit efficient search for dirty-paper code words. These methods are discussed in this section.

2.6.1 Dirty paper trellis watermarking

A *trellis code* (or *convolutional code*) is an error correction code (ECC). For its details and theory, readers are directed to [Vit95].

Trellis coding was exploited by Miller *et al.* [MDC02, MDC04] who proposed dirty paper trellis coding for watermarking. Their dirty paper trellis for watermarking has two differences compared to traditional trellis. Firstly, unlike the one-to-one mapping of messages to code words in a traditional trellis, multiple code words exist to encode the same message in a dirty paper trellis. Secondly, each code word generated by the dirty paper trellis for watermarking is a real valued vector, instead of a binary bits sequence in the traditional trellis.

Efficient detection of the dirty paper trellis watermark is accomplished by exploiting a Viterbi decoder. The Viterbi algorithm finds the closest code word to the watermarked Work. This is measured by the *Euclidean distances* between the vector of watermarked Work and the vectors presenting the code words.

All code words generated by the dirty paper trellis are required to have the same energy, i.e., are equi-energetic. They are all considered to be placed on the surface of a multidimensional sphere. One benefit of these *spherical codes* is that the detection for a code word is unaffected by amplitude scaling of the watermarked signal. This is illustrated with the circle (two-dimensional sphere) in Figure 2.4. While it is a simple example, the explanations can be extended to N-dimensional space.

Here in Figure 2.4, the code words, $\mathbf{w_0}, \mathbf{w_1}, \cdots, \mathbf{w_n}$, are distributed on the circle. All vectors that share the same closest code word $\mathbf{w_0}$ are in the shaded region. This shaded region is often referred to as the *detection region* of $\mathbf{w_0}$. Because all spherical code words are equi-energetic, the direction, not the magnitude, of a given vector determines which code word is the closest one to the vector. In other words, the question of which code word's detection region the signal vector belongs to is determined by the vector's direction. Consider a vector $\mathbf{y}$ which is in the detection region of $\mathbf{w_0}$, and becomes $\beta\mathbf{y}$ after its amplitude is scaled by a factor of β. This scaling only changes the magnitude, *not* its direction. Consequently, the scaled version $\beta\mathbf{y}$ remains in the same

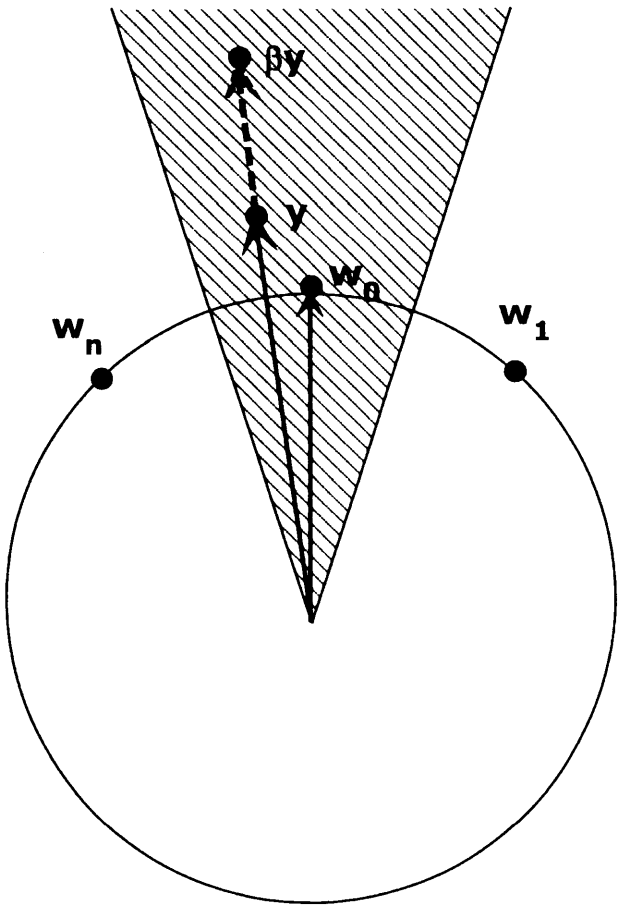


Figure 2.4: Detection of a spherical code is unaffected by amplitude scaling of the signal.

detection region. This is why the detection of spherical codes is invariant to amplitude scaling.

Though fairly robust to several signal processes, Miller *et al.*'s method [MDC04] can introduce distortion. Another concern for dirty paper trellis watermarking is its computational complexity. A particular informed embedding strategy is proposed [MDC04] to ensure that a given level of robustness is obtained with a tolerable distortion level. However, while selection of the dirty paper code is efficient, subsequent informed embedding of the code word requires a computational expensive iterative procedure. According to our experiences, it takes over ten minutes to do informed coding and informed embedding on a 240×368 image using an Intel 1.7 GHz Pentium 4 PC.

There are several papers offering interesting extensions to dirty paper trellis wa-

termarking [WMC04, APGM05]. A method is proposed in which the best path is found according to correlation or a combination of correlation with a perceptual model [WMC04]. Abrardo *et al.* [APGM05] recently reported work on a combination of dirty paper trellis coding and rational dither modulation (RDM). RDM is a lattice based watermarking scheme which exploits adaptive quantization step sizes at both embedder and decoder. The main purpose of RDM is to provide invariance to amplitude scaling. We improve on RDM by using a modified perceptual model discussed in Chapter 4.

2.6.2 Lattice code watermarking

Lattice based watermarking algorithms, more often referred to as quantization index modulation (QIM), where introduced by Chen and Wornel [BG99a, BG99b, CW00, BG01b, BG01a] for data hiding. In this scheme, a set of vectors are extracted from the host signal. They are then quantized by means of a quantizer chosen from a pool of predefined quantizers on the basis of the watermark message to be embedded. Note that the traditional QIM method uses a set of uniform scalar quantizers.

QIM methods are appealing to watermarking researchers due to their easy implementation, good computational efficiency and considerable capacity. But they also have some limitations. Firstly, the distortion due to embedding is not adaptive and may be perceptually noticeable. Secondly, unlike the spherical codes generated by the dirty paper trellis, lattice based QIM methods are extremely sensitive to amplitude scaling. Thirdly, they are vulnerable to re-quantization noise, for example, JPEG compression.

Details of standard QIM methods are further discussed in Chapter 3. In order to take advantages of QIM's computational efficiency over dirty paper trellis watermarking, we attempt to overcome the limitations of QIM in this thesis. Our methods and experimental results are mainly described and illustrated in Chapters 4 and 5.

2.7 Importance of Perceptual Models

The last twenty years have witnessed a rapid adoption of digital media formats, e.g. CD, JPEG, MP3, DVD. A number of factors contributed to this trend, including low cost digital devices, increased computing power, and more bandwidth. However, even with

increased bandwidth, there has been a need for compression algorithms that greatly reduce the size of the data, yet maintain high quality.

Consider audio CD as an example which does not use compression. Compact discs sample audio at 44.1 kilohertz (kHz), using pulse code modulation (PCM) with 16-bits-per-sample resolution. Thus, CDs need a date rate of 1.41 Mbps for a stereo channel. Unfortunately, there are many application scenarios, like wireless networks, where constraints such as channel bandwidth, storage capacity, and low cost prohibit such high data rates. In these cases, it is necessary to deliver high quality digital contents at low bit rates. In response to this need, lossy perceptual coding methods have been developed for audio, images and video.

Human perceptual models are investigated to help design compression methods which can produce high quality but low data rate digital contents. It is realized that (i) natural sounds / images contain large amounts of redundant information, (ii) human perception is not equally sensitive to all information present in audio and visual signals. Consequently, human perceptual models are often analytically and experimentally devised to understand where and by how much content can be compressed such that the changes remain imperceptible.

Fidelity and quality are frequently used to evaluate multimedia content. Fidelity refers to the similarity of content before and after processing. High fidelity indicates that there is little or no perceptual differences (distortions) between the original content and the processed one. Conversely, a low fidelity copy is quite distinguishable from the original content. Fidelity can only be characterised when the original and the processed copy are available for comparison. In contrast, quality is an absolute measure of a digital content, and no comparison with other content is needed. For example, the quality of an image or audio clip is simply measured by how good it looks or sounds which depends on various parameters such as sampling frequency, resolution, compression rate, etc.

The difference between fidelity and quality needs to be clarified: a high fidelity (or even identical) copy could be of low quality simply because of the low quality original

Work. On the other hand, it is also possible to have high quality but low fidelity. An image may be retouched using software like PhotoShop to improve the quality of an image but it may have low fidelity if the difference between the processed and original image is significant. More discussions about the relationship between fidelity and quality can be found in [SF96]. However, in digital watermarking we are primarily interested in high fidelity rather than high quality.

Perceptual models and perceptual measurement are relevant and important to digital watermarking. Perceptual models can help to improve watermark imperceptibility by increasing fidelity. Measures of perceptual distortion allow us to assess the impact of a particular watermarking algorithm. The mean square error (MSE), defined in Equation (2.1), is one popular way to measure the differences between multimedia signals. Although it is often used, MSE can not provide reliable predictions of perceptual distortions [Gir93, WB06]. The reason is that the MSE function does not take into account human perceptual properties, unlike perceptual models.

An early and simple perceptual model is based on the observation by Ernst Weber, a 19th century experimental psychologist, that the smallest perceptible change to an image, i.e., the so called just noticeable difference (JND), is proportional to the background stimulus value. This is arguably the first and best known perceptual model, which can be simply expressed as follows:

$$\frac{\delta x}{x} = T$$

where δx is the JND, x is the initial stimulus and T is a constant value depending on various environments.

Weber's law indicates that human eyes are more sensitive to the noise added to a dark image (with low pixel values) than the noise added to a bright image (with high pixel values). This is demonstrated by the examples in Figures 2.5 and 2.6. Figure 2.5(a) is a generally bright image, named "river", and Figure 2.5(b) is its distorted version. The MSE between this pair of images is 30. The original image of "buildings",

shown in Figure 2.6(a), is much darker compared with the original "river". Figure 2.6(b) is a distorted "buildings" image. The MSE between the pair of "buildings" images is also 30. Although the MSE measurements are identical, the distortion between the pair of dark "buildings" images is perceptually much higher than the distortion between the pair of bright "river" images.

2.8 Structural Similarity (SSIM) Model

Mean square error (MSE) and peak signal-to-noise ratio (PSNR) are widely used to evaluate distortion or fidelity between images. This is because they are easy to understand and simple to calculate. Nevertheless, as discussed in previous section, they do not match well with human perception. Many efforts have been taken to develop perceptual models based on known characteristics of human visual system (HVS).

Wang *et al.* [WBSS04] proposed a model to perceptually and quantitatively assess differences between images. This structural similarity model, SSIM as it is called, is based on the hypothesis that the HVS is highly adapted for extracting structural information.

Given two image signals x and y (y could be a distorted, compressed, reconstructed or watermarked version of x), the similarity between them can be measured by the SSIM model, which serves as a quantitative metric for perceptual differences between images. Figure 2.7 shows the system diagram for this structural similarity (SSIM) measurement model.

The SSIM system computes a similarity measurement by doing three comparisons: *luminance, contrast* and *structure*. Note that a *local window* is needed to compute parameters for *local* luminance and contrast because luminance and contrast vary according to local features of an image. The structural information of an image, defined in [WBSS04], is the attributes representing the structure of objects but being independent of the average luminance and contrast.

The similarity measurement functions, denoted as $SSIM(x, y)$, is required to satisfy the following properties:

(a) The original "river" image

(b) The distorted "river" image

Figure 2.5: A bright "river" image and its distorted version with MSE=30

(a) The original "buildings" image

(b) The distorted "buildings" image

Figure 2.6: A dark image "buildings" and its distorted version with MSE=30

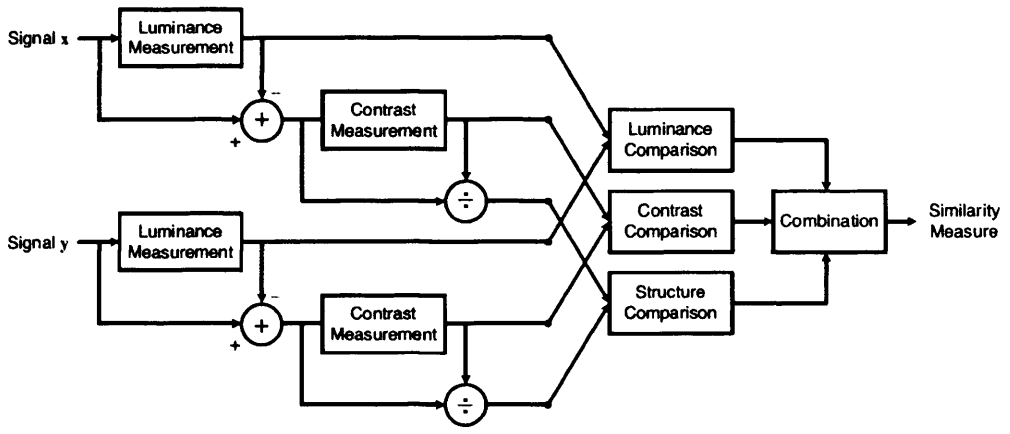


Figure 2.7: Structural similarity (SSIM) measurement system, taken from [WBSS04].

(1) Symmetry: $SSIM(x, y) = SSIM(y, x)$;

(2) Boundeness: $SSIM(x, y) \leq 1$;

(3) Unique maximum: $SSIM(x, y) = 1$; if and only if $x = y$, which means signal x and y are identical.

These constraints need to be considered when constructing a $SSIM(x, y)$ function. Based on the diagram shown in Figure 2.7, the $SSIM(x, y)$ is given by:

$$SSIM(x, y) = f(lum(x, y), con(x, y), str(x, y)) \tag{2.7}$$

The overall similarity measurement function, $SSIM(\cdot)$, is a combination of $lum(x, y), con(x, y)$ and $str(x, y)$, which are functions of *luminance comparison, contrast comparison* and *structure comparison*, respectively. Details of these three functions are discussed as follows:

Luminance comparison function

Given N samples of an image signal, $x_i, i = 1, 2, ..., N$, the average of luminance intensity, μ_x, is calculated as:

$$\mu_x = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{2.8}$$

For another image y, the average μ_y is also calculated. The $lum(x, y)$ is then a function of μ_x and μ_y, given by:

$$lum(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{2.9}$$

where C_1 is a constant which is used to avoid instability when $\mu_x^2 + \mu_y^2$ tend to zero. In [WBSS04], authors specifically suggest it to be:

$$C_1 = (K_1 L)^2 \tag{2.10}$$

where $K_1 \ll 1$ is a small constant number and L is the dynamic range of the pixel values, for example, $L = 255$ for 8-bit-per-pixel grayscale images. Note that equation (2.9) satisfies the three properties mentioned above: symmetry, boundeness and unique maximum.

Contrast comparison function

In order to compare contrast comparison metric, the average intensity of the signal needs to be removed. For a discrete signal, the resulting signal x' corresponding to x is given by $x' = x - \mu_x$, it consequently has the feature that $\sum_{i=1}^{N} x_i' = 0$. The standard deviation of x is used as an estimate of the signal contrast, which is given by:

$$\sigma_x = \left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2\right)^{\frac{1}{2}} \tag{2.11}$$

The contrast comparison $con(x,y)$ is then a function of σ_x and σ_y, formulated as:

$$con(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{2.12}$$

where $C_2 = (K_2 L)^2$ and $K_2 \ll 1$. This function again satisfies the three properties mentioned earlier in this section. The change of contrast of the two signals is defined as $\delta\sigma = \sigma_y - \sigma_x$. Note that given the same amount of $\delta\sigma$, the value of the function $con(x,y)$ is smaller for higher contrast σ_x than for lower σ_x. This is consistent with the contrast masking feature of human visual system (HVS) which indicates that changes to highly textured areas of an image are less perceptually noticeable than in plain area.

Structural comparison function

The two image signals to be compared need to be normalized by their own corresponding standard deviation, so that they have unit standard deviation. Structure comparison is computed after luminance subtraction and variance normalization. Specifically, the structure comparison function $str(x, y)$ is calculated with normalized signals $(x - \mu_x)/\sigma_x$ and $(y - \mu_y)/\sigma_y$. The structural similarity is then the correlation (inner product) between those two normalized signals. Note that the inner product between $(x - \mu_x)/\sigma_x$ and $(y - \mu_y)/\sigma_y$ is equivalent to the correlation between x and y. The structural comparison function is then given by:

$$str(x, y) = \frac{\mu_{xy} + C_3}{\mu_x \mu_y + C_3} \tag{2.13}$$

where C_3 is a small constant and μ_{xy} is defined as:

$$\mu_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y) \tag{2.14}$$

SSIM function

SSIM is finally formed as a combination of the three comparisons shown in Equations (2.9), (2.12) and (2.13).

$$SSIM(x, y) = [lum(x, y)]^{\alpha} [con(x, y)]^{\beta} [str(x, y))]^{\gamma} \tag{2.15}$$

where $\alpha > 0, \beta > 0$ and $\gamma > 0$, and these three sub-functions are taken to adjust the relative importance between one another. It is easily seen that $SSIM(x, y)$ function satisfies the three desired properties mentioned earlier in this section.

In order to obtain a simplified version of SSIM, the parameters can be set as $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$. This results in a specific formulation of SSIM, given by

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\mu_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\mu_x^2 + \mu_y^2 + C_2)} \tag{2.16}$$

2.9 Stochastic Approaching Model

Voloshynovskiy *et al.* [VHBP99] proposed another perceptual model for digital image watermarking. It can be used in transform domains like Fourier, DCT or Wavelet domains as a perceptual modulation function.

The fundamental idea relies on an appropriate stochastic modeling of the original host signal. Once the stochastic models of the host signal and the watermarked signal are determined, capacity of the image to watermarking can be estimated and the detection problem can be formulated based on a Bayesian framework.

For further discussions, the watermarking problem is modeled as $\mathbf{y} = \mathbf{x}+\mathbf{w}$, where $\mathbf{x}$ is the host signal, $\mathbf{y}$ is the generated watermarked signal and $\mathbf{w}$ is the watermark.

Noise visibility function (Texture masking function)

As noted earlier, modifications to an image are more perceptible in plain areas than in highly textured areas. Based on this fact, the texture masking function is related to the noise visibility function (NVF) in [VHBP99], which is formulated as:

$$NVF = \frac{\varpi\sigma_w^2}{\varpi\sigma_w^2 + \sigma_x^2}, \tag{2.17}$$

where σ_x and σ_w are the standard deviation of $\mathbf{x}$ and $\mathbf{w}$, respectively. The weighting, ϖ, is used to determine the particularities of the perceptual model, and is discussed later in this section. The NVF is an output of the perceptual model which indicates how a noise is perceived, It can also indicate the level of image smoothing.

For instance, in plain regions, samples of host signal tend to be uniform so that $\sigma_x^2 \rightarrow 0$, which consequently results in $NVF \rightarrow 1$. In contrast, in highly textured regions, $\sigma_x^2 \gg \sigma_w^2$, thus $NVF \rightarrow 0$, where it is considerably insensitive to the noise added. Two particular cases of NVF will be discussed next.

NVF based on non-stationary Gaussian model

This model assumes the host signal, e.g., an image, is randomly distributed with non-stationary mean. That is, a host signal is assumed to be a locally $i.i.d.$. In this case,

the NVF function is computed by using the non-stationary variance according to the quadratic energy function. This is shown as follows:

$$NVF(i,j) = \frac{1}{1+\sigma_x^2(i,j)}, \tag{2.18}$$

Where σ_x is the local standard deviation of the partial image centered on the pixel with coordinates (i,j). Thus, the NVF is roughly inversely proportional to the local image energy, defined by the variance. Assuming the image is a locally $i.i.d.$ Gaussian, the estimation of the local image variance is formulated as:

$$\sigma_x^2(i,j) = \frac{1}{(2L+1)^2} \sum_{k=-L}^{L} \sum_{l=-L}^{L} (x(i+k,j+l) - \bar{x}(i,j))^2 \tag{2.19}$$

where the window size is $(2L+1) \times (2L+1)$ and the average $\bar{x}$ is computed as:

$$\bar{x}(i,j) = \frac{1}{(2L+1)} \sum_{k=-L}^{L} \sum_{l=-L}^{L} x(i+k,j+l). \tag{2.20}$$

NVF based on stationary Generalized Gaussian model

In contrast to the non-stationary Gaussian model, the stationary generalized Gausian (GG) model assumes the host signal to be globally $i.i.d.$. In this case, the NVF is formulated as:

$$NVF(i,j) = \frac{\varpi(i,j)}{\varpi(i,j)+\sigma_x^2}, \tag{2.21}$$

where

$$\varpi(i,j) = \gamma[\eta(\gamma)]^{\gamma} \frac{1}{\|r(i,j)\|^{2-\gamma}},$$

$$r(i,j) = \frac{x(i,j) - \bar{x}(i,j)}{\sigma_x},$$

$\eta(\gamma) = \sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}}$, and $\Gamma(t) = \int_0^\infty e^{-u}u^{t-1}du$. The parameter γ is the shape parameter, which is in the range of $0.3 \leq \gamma \leq 1$ for most real images.

Using this perceptual model to determine the noise visual function (NVF), a con-

tent adaptive watermarking scheme is formulated as:

$$y = x + (1 - NVF) \cdot w_m \cdot \alpha$$

where w_m is the watermarking message mark and α is the embedding strength.

The proposed watermarking scheme based on this perceptual model embeds the watermark more strongly in highly textured regions than in plain areas. However, this model does not consider the luminance sensitivity of the human visual system.

2.10 Watson's Perceptual Model

Any perceptual model of the human visual system (HVS) has to account for a variety of perceptual phenomena, including luminance masking, contrast masking and sensitivity, all of which are discussed shortly. In psychophysical studies, the level of distortion that can be perceived in just over 50% of experimental trials is often referred to as a *just noticeable difference*, or JND. This difference is considered the minimum that is generally perceptible and JNDs are sometimes employed as a unit for measuring the distance between two images.

Watson's model estimates the perceptibility of changes in individual terms of an image's block DCT[1]. It uses the block DCT transform, which proceeds by first dividing the image into disjoint 8×8 blocks of pixels. Each of these blocks is then transformed into the DCT domain, resulting in the block DCT coefficients of the transformed image, C. We denote one term of the $k - th$ block by $C[i, j, k], 0 \leq i, j \leq 7$. $C[0, 0, k]$ is the DC term, i.e. the mean pixel intensity in the block.

Watson's model consists of a sensitivity function, two masking components based on luminance and contrast masking, and a pooling component.

Sensitivity

The model defines a frequency sensitivity table, t. For each DCT coefficient, (i, j), each table entry, $t[i, j]$, is approximately the smallest magnitude of discernible change

[1]Note that we are not referring to quantized JPEG coefficients.

in the absence of any masking noise. The resulting frequency sensitivity table is shown in Table 2.1. Note that it is a table of constant values.

1.40	1.01	1.16	1.66	2.40	3.43	4.79	6.56
1.01	1.45	1.32	1.52	2.00	2.71	3.67	4.93
1.16	1.32	2.24	2.59	2.98	3.64	4.60	5.88
1.66	1.52	2.59	3.77	4.55	5.30	6.28	7.60
2.40	2.00	2.98	4.55	6.15	7.46	8.71	10.17
3.43	2.71	6.34	5.30	7.46	9.62	11.58	13.51
4.79	3.67	4.60	6.28	8.71	11.58	14.50	17.29
6.56	4.93	5.88	7.60	10.17	13.51	17.29	21.15

Table 2.1: DCT frequency sensitivity table.

Luminance Masking

Luminance adaptation refers to the fact that a DCT coefficient can be changed by a larger amount before becoming perceptible, if the average intensity of the 8×8 block is brighter. The luminance-masked threshold, $t_L[i, j, k]$, is given by

$$t_L[i, j, k] = t[i, j](C_o[0, 0, k]/C_{0,0})^{\alpha_T} \tag{2.22}$$

where α_T is a constant with a suggested value of 0.649, $C_o[0, 0, k]$ is the DC coefficient of the kth block in the original image, and $C_{0,0}$ is the average intensity of the image. Alternatively, $C_{0,0}$ may be set to a constant value representing the expected intensity of images.

Contrast Masking

Contrast masking, i.e., the reduction in visibility of a change in one frequency due to the energy present in that frequency, results in a masking threshold, $S[i, j, k]$, given by

$$S[i, j, k] = \max\left(t_L[i, j, k], |C_o[i, j, k]|^{0.7} t_L[i, j, k]^{0.3}\right) \tag{2.23}$$

The final threshold,$S[i,j,k]$, estimates the amounts by which individual terms of the block DCT may be changed before resulting in one JND. We refer to these thresholds as *slack*.

Watson Distance

Given the original content, c_o, its transform domain coefficients are denoted by C_o, and C_w denotes the watermarked transform coefficients. The Watson Distance between the original and watermarked image is then given by:

$$D_{wat}(c_o, c_w) = \left\{ \sum_{i,j,k} \left(\frac{C_w[i,j,k] - C_o[i,j,k]}{S[i,j,k]} \right)^4 \right\}^{\frac{1}{4}} \tag{2.24}$$

As an aside, we note that the JPEG standard provides an example set of image-independent quantization matrices, which, while not part of the standard, have been very widely adopted. The JPEG standard requires that the quantization matrix be part of the compressed file. Watson's model can be used to determine an image-dependent quantization matrix that can significantly improve the performance of JPEG. The interested reader is directed to [Wat93b] for further information.

To get better compression methods for images, people need to find optimal balance point between higher compression rate and lower perceptual artifact. Similarly, to obtain better digital watermarking methods, it is required to develop more elaborate compromise between robustness and fidelity of watermarked images. For both demands, it is very important and desirable to investigate visual perceptual models. The Watson's model introduced in this section is used later to design watermarking schemes.

Chapter 3

Quantization Index Modulation

Quantization index modulation (QIM), first proposed by Chen and Wornell [BG99b], provides a computationally efficient method for implementing codes based on Costa's work [Cos83]. QIM uses a structured lattice code to provide a computational efficient watermarking algorithm with high data capacity. Chen and Wornel introduce a class of watermarking methods called quantization index modulation (QIM) [BG99a,BG99b,CW00,BG01b,BG01a]. This class of techniques embed the watermark in a host signal through quantization; a different quantization vector is used to embed a different watermark message. This chapter describes in detail QIM and its variants such as dither modulation (DM), distortion-compensated QIM (DC-QIM), spread transform dither modulation (STDM). Some limitations of QIM and previous works addressing these issues are also discussed in the later part of this chapter.

3.1 Basics of Quantization Index Modulation

The basic function that QIM uses is a quantizer. It maps a value to the nearest point belonging to a class of pre-defined discontinuous points. The standard quantization operation with step size Δ is defined as

$$Q(x, \Delta) = \Delta round\left(\frac{x}{\Delta}\right) \tag{3.1}$$

where x is a signal sample, Δ is the quantization step size, and the function $round(\cdot)$ denotes rounding a value to the nearest integer.

Basic quantization index modulation uses two quantizers, Q_0 and Q_1. They can be used to quantize the host signal to two sets of disjoint points, one set represents bit '0' while the other represents bit '1'. When a watermarking message is being embedded, Q_0 or Q_1 is chosen according to the message bit to quantize x into the nearest quantization point. For example, Q_0 and Q_1 may be chosen such that Q_0 quantizes to even integers and Q_1 quantizes to odd integers. This scheme is illustrated in Figure 3.1, in which the set of circles, $\circ$, indicate lattice points for quantizer Q_0 which represents bit '0' and the set of crossings, $\times$, indicate lattice points for the quantizer Q_1 which represents bit '1'.

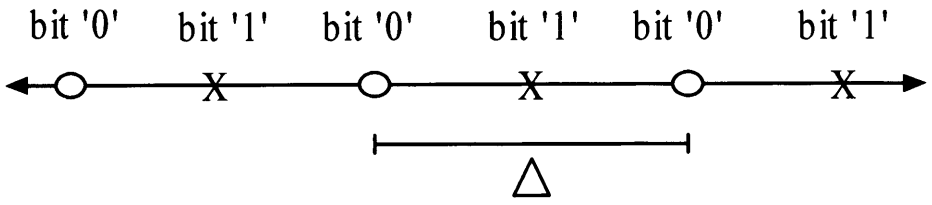


Figure 3.1: Basic quantization index modulation

If we wish to embed a '0'-bit, then Q_0 is chosen, otherwise Q_1 is used. Given a particular binary bit b from a message $\mathbf{m}$ ($b \in \mathbf{m} = \{0, 1\}$), the watermarked signal y is given by:

$$y = Q_b(x, \Delta_b) \tag{3.2}$$

Normally, these two quantizers Q_0 and Q_1 share an equal step size. Assuming that $\Delta_0 = \Delta_1 = \Delta$, and one quantizer can be obtained by simply shifting the other one by $\Delta/2$. In this case, the basic QIM method can also be formulated with one quantization function as follows:

$$\mathbf{y} = Q(\mathbf{x}, \mathbf{m}, \Delta) = \Delta round(\frac{\mathbf{x}}{\Delta} - \frac{\mathbf{m}}{2}) + \frac{\mathbf{m}}{2}\Delta \tag{3.3}$$

where $\mathbf{x}$ is the host signal vector, $\mathbf{m}$ is the message vector, Δ is quantization step size and $\mathbf{y}$ is the watermarked signal vector. Note that the codebook of dither modulation is entirely defined by the step size Δ (and possibly a shift). This significantly reduces the

amount of memory needed to store the infinity of codewords used here.

3.1.1 Minimum distance between codewords

The minimum distance, d_{min}, is the minimum distance between any two generated codewords. In the case of QIM, given the conditions described in Equation (3.3) and Figure 3.1, the minimum distance is:

$$d_{min} = \frac{1}{2}\Delta \tag{3.4}$$

This minimum distance, d_{min}, plays an important role as it effectively determines the robustness of the coding to noise. The minimum distance d_{min} also indicates how much perturbation can be tolerated by the system. For example, intuitively, a perturbation of over half d_{min}, i.e. $1/4\Delta$ for basic QIM, added to watermarked signal may cause a detection error.

3.1.2 Embedding-induced distortion

Embedding-induced distortion refers to the distortion introduced to the original host signal by the watermarking embedding process, that is, the difference between the original signal and the watermarked signal. We can measure the embedding-induced distortion, denoted as D_e, by the expectation of the mean square error between the host signal and the watermarked signal. This measurement for QIM is:

$$\begin{aligned} D_e &= E(MSE_e) \\ &= E\left((Q(\mathbf{x}, \mathbf{m}, \Delta) - \mathbf{x})^2\right) \qquad (3.5) \\ &= E\left(\frac{1}{L}\sum_{i=1}^{L}(y_i - x_i)^2\right) \end{aligned}$$

(3.6)

where L is the length of the signal vector $\mathbf{x}$, $\mathbf{m}$ and $\mathbf{y}$, that is, the total number of signal samples to be measured. x_i and y_i is a sample of original signal and watermarked signal, respectively.

If we define the difference between the watermarked signal $\mathbf{y}$ and the original signal $\mathbf{x}$ as watermark: $\mathbf{w} = \mathbf{y} - \mathbf{x}$, the document-to-watermark ratio is defined by:

$$DWR = 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_w^2}\right) \tag{3.7}$$

where σ_x^2 and σ_w^2 are the variance of $\mathbf{x}$ and $\mathbf{w}$, respectively. In the high Document-to-Watermark ratio regime of primary interest for high-fidelity watermarking applications, we can assume that the quantization cells are sufficiently small that the host signal can be modeled as uniformly distributed within each cell. We also assume that the message to be embedded is independent of the host signal.

In this case, the embedding-induced distortion is:

$$\begin{aligned} D_e &= E\left((Q(\mathbf{x},\mathbf{m},\Delta)-\mathbf{x})^2\right) \\ &= \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} x^2\,dx \\ &= \frac{1}{12}\Delta^2 \end{aligned} \tag{3.8}$$

Thus the embedding-induced distortion for standard QIM is $\frac{1}{12}\Delta^2$, which is only a function of quantization step size.

3.1.3 QIM detector: hard and soft decision

For a QIM detector, the received signal, $\mathbf{r}$, is obtained after the watermarked signal, $\mathbf{y}$, undergoes distortion. This distortion is modeled as a random noise process, although many distortions are not random. The received signal is therefore given by $\mathbf{r} = \mathbf{y}+\mathbf{v} = \mathbf{x}+\mathbf{w}+\mathbf{v}$, where $\mathbf{v}$ is an unknown noise source. For a blind detector, the QIM detector has no knowledge of $\mathbf{y}$ or $\mathbf{x}$, but assumes that $\mathbf{r}$ and $\mathbf{y}$ are in the same quantization bin (detection region), otherwise a detection error may occur. Based on this assumption, the QIM detector works as a minimum-distance decoder: First, $\mathbf{r}$ is used to reconstruct two possible watermarked output points to represent bit '0' and bit '1' respectively, which is equivalent to embedding '0' and '1' into $\mathbf{r}$ in the same way as the the QIM embedder does. The estimated message bit, $\hat{b}$, is then detected by judging which of

these two reconstructed points has the minimum Euclidean distance to the received signal $\mathbf{r}$. It is formulated as:

$$\hat{b} = \underbrace{argmin}_{b\in 0,1} (r - Q(r, b, \Delta))^2 \tag{3.9}$$

The above description embedded one bit in each sample. In practice, we usually spread one message bit into a sequence of N samples. One way to achieve this is to use a rate $1/N$ repetition encoding, i.e. simply embed the same message bit in N samples. Detection can still be performed on a one bit per sample basis followed by a majority vote taken over the N samples to decide which message bit was embedded. We refer to this as *hard decision* detection, and is given by:

$$\begin{aligned} &\hat{m_n} = \lfloor \tfrac{2}{N+1} \textstyle\sum_{h=(n-1)N+1}^{nN} \underbrace{argmin}_{b\in 0,1} (r_h - Q(r_h, b, \Delta))^2 \rfloor, \\ &n = 1, 2, \ldots, L/N, \ \lfloor\ \rfloor \text{ is the floor function} \end{aligned} \tag{3.10}$$

where L is the length of the vector, r_h is the $h-th$ element of the vector $\mathbf{r}$ and $\hat{m_n}$ is n-th message bit.

An alternative detection strategy is to accumulate the two Euclidean distances for N samples and then determine the detected message bit, i.e.,

$$\begin{aligned} &\hat{m_n} = \underbrace{argmin}_{b\in 0,1} \textstyle\sum_{h=(n-1)N+1}^{nN} (r_h - Q(r_h, b, \Delta))^2, \\ &n = 1, 2, \ldots, L/N. \end{aligned} \tag{3.11}$$

The code rate is also $1/N$ but this *soft decision* decoding usually outperforms hard decision decoding.

3.2 Dither Modulation

Dither Modulation [BG99a] is one implementation of QIM, in which a dither signal is used. Dither quantization is a modified quantization process sometimes used to control the effect of quantization noise. A dither signal is added to the original host signal,

prior to quantization. This dither signal is usually a pseudo-random signal. After quantization, the signal is commonly transmitted to one or more receivers that then attempt to reconstruct the signal. Schuchman [Sch64] showed that if a well-chosen dither signal is added to the input signal prior to quantization, the quantization error will be independent of the input host signal (generally original images, audio and video for watermarking systems). Note that dither modulation has no impact on the average distortion which remains equal to $\Delta^2/12$.

The purpose of dither modulation in QIM [RM96, NP84] is threefold. First, it is well recognized that a pseudo-random dither signal can reduce quantization artifacts to produce a perceptually superior quantized signal. Second, dither ensures that the quantization noise is independent of the host signal, $\mathbf{x}$. The output values of $\mathbf{y}$ are now uniformly distributed along the real axis. Third, the pseudo-random dither signal can be considered to act as a key which is only known to the watermark embedder and detector, thereby improving the security of the system. Without access to the key, the detector is unable to re-generate the dither signal and can not extract the hidden message.

The host signal sample x_n is quantized with the resulting dithered quantizer to form the watermarked signal sample y_n. The embedding function embeds message bit m_n by:

$$y_n(x_n, m_n) = Q(x_n + d(n, m_n), \Delta) - d(n, m_n). \tag{3.12}$$

where

$$d[n,1] = \begin{cases} d[n,0] + \Delta/2, & d[n,0] < 0. \\ d[n,0] - \Delta/2, & d[n,0] > 0. \end{cases} \quad n = 1, 2, \ldots, L. \tag{3.13}$$

and $d(n,0)$ is a pseudo-random signal usually chosen with a uniform distribution over $[-\Delta/2, \Delta/2]$ and L is the total number of signal samples.

3.2.1 DM detector

During detection, the detector calculates two signals, $s(n,0)$ and $s(n,1)$ by embedding '0' and '1'into the received signal $\mathbf{r}$ separately, in the same manner as Equation (3.12).

$$\begin{aligned} s(n,0) &= Q(r_n + d[n,0], \Delta) - d[n,0]; \\ s(n,1) &= Q(r_n + d[n,1], \Delta) - d[n,1]. \end{aligned} \tag{3.14}$$

The detected message bit is then determined by judging which of these two signals has the minimum Euclidean distance to the received signal $\mathbf{r}$

$$\hat{m_n} = \underbrace{argmin}_{b\in 0,1}(r_n - s(n,b))^2 \tag{3.15}$$

The above description embedded one bit in each sample. Again, if we would spread one message bit into a sequence of N samples, a rate $1/N$ repetition encoding can be exploited, i.e. simply embed the same message bit in N samples. Similar with discussions in Section 3.1.3, we have two options to do detection: hard or soft decision. The hard decision detector is given by:

$$\begin{aligned} &\hat{m_n} = \lfloor \tfrac{2}{N+1} \textstyle\sum_{h=(n-1)N+1}^{nN} \underbrace{argmin}_{b\in 0,1}(r_h - s(h,b))^2 \rfloor, \\ &n = 1, 2, \ldots, L/N, \text{ where } \lfloor\ \rfloor \text{ is the floor function} \end{aligned} \tag{3.16}$$

Alternatively, with the same code rate, the soft decision detector is:

$$\begin{aligned} &\hat{m_n} = \underbrace{argmin}_{b\in 0,1} \textstyle\sum_{h=(n-1)N+1}^{nN} (r_h - s(h,b))^2, \\ &n = 1, 2, \ldots, L/N. \end{aligned} \tag{3.17}$$

The code rate is also $1/N$ but this *soft decision* decoding [BG01b] usually outperforms hard decision decoding.

3.3 Distortion-compensated QIM

In [BG01a] Chen *et al.* introduced distortion compensated QIM (DC-QIM), which can find optimal balance point between robustness and capacity under a certain additive white Gaussian noise (AWGN). Furthermore, the distortion compensated version can also be exploited to DM, yielding distortion compensated dither modulation (DC-DM).

For distortion compensated QIM, the watermarked signal is obtained by adding back part of the quantization error, controlled by distortion compensation parameter α, to the quantized host signal. According to Chen's paper [BG01a], The embedding equation for DC-QIM is:

$$\mathbf{y} = Q(\mathbf{x}, \mathbf{m}, \Delta/\alpha) + (1-\alpha)\left(\mathbf{x} - Q(\mathbf{x}, \mathbf{m}, \Delta/\alpha)\right) \tag{3.18}$$

It can be reformed as:

$$\begin{aligned} \mathbf{y} &= \mathbf{x} + \alpha\left(Q(\mathbf{x}, \mathbf{m}, \Delta/\alpha) - \mathbf{x}\right) && (3.19)\\ &= \alpha Q(\mathbf{x}, \mathbf{m}, \Delta/\alpha) + (1-\alpha)\mathbf{x} && (3.20) \end{aligned}$$

Since

$$Q(\mathbf{x}, \mathbf{m}, \Delta/\alpha) = \frac{1}{\alpha} Q(\alpha\mathbf{x}, \mathbf{m}, \Delta) \tag{3.21}$$

Then

$$\mathbf{y} = Q(\alpha\mathbf{x}, \mathbf{m}, \Delta) + (1-\alpha)\mathbf{x} \tag{3.22}$$

Consequently, the expected embedding induced square error is:

$$\begin{aligned} D_e &= E\left((\mathbf{y}-\mathbf{x})^2\right) \\ &= E\left(\alpha^2\left(Q(\mathbf{x}, \mathbf{m}, \Delta/\alpha) - \mathbf{x}\right)^2\right) \\ &= \alpha^2 E\left(Q(\mathbf{x}, \mathbf{m}, \Delta/\alpha) - \mathbf{x}\right)^2 \\ &= \alpha^2 \frac{1}{12}\left(\frac{\Delta}{\alpha}\right)^2 \end{aligned}$$

$$= \frac{1}{12}\Delta^2 \tag{3.23}$$

Thus, the expectation of embedding-induced square error distortion, D_e, for DC-QIM is same with QIM discussed in Section 3.1.2. In fact, as shown in the second term of Equation (3.18), i.e., $(1-\alpha)\,(\mathbf{x} - Q(\mathbf{x}, \mathbf{m}, \Delta/\alpha))$, a fraction of $(1-\alpha)$ of the quantization error is added back in order to maintain the same embedding-induced distortion. The distortion compensated factor α controls the compensation for the distortion.

Although DC-QIM maintains the same embedding-induced distortion. It does have an impact on the minimum distance between constructed code words. We can see from Section 3.1.1 that the minimum distance is a function of step size used for the quantization. For DC-QIM the step size is influenced by the distortion compensated factor α, which is inversely proportional to the minimum distance. Thus decreasing α leads to greater minimum distance, but still maintains the same embedding-induced distortion by adding back the distortion-compensated interference. If we consider the quantization power of quantization as signal power and both the distortion compensation part and channel noise as interference, then the signal-to-noise ratio can be written as:

$$\begin{aligned} SNR &= \frac{\frac{D_e}{\alpha^2}}{(1-\alpha)^2\frac{D_e}{\alpha^2} + \sigma_v^2} \\ &= \frac{D_e}{(1-\alpha)^2 D_e + \alpha^2\sigma_v^2} \end{aligned} \tag{3.24}$$

where σ_v^2 is the variance of the noise in the channel.

To maximize the SNR, the optimal parameter α should be:

$$\begin{aligned} \alpha_{opt} &= \frac{D_e}{D_e + \sigma_v^2} \\ &= \frac{DNR}{DNR+1} \end{aligned} \tag{3.25}$$

where DNR is the embedding-induced distortion-to-noise ratio which is defined

as:

$$DNR = \frac{D_e}{\sigma_v^2} \tag{3.26}$$

3.3.1 DC-QIM detector

For DC-QIM, given a received signal $\mathbf{r}$, bit '0' and '1' are embedded into $\mathbf{r}$ (in the same way as the DC-QIM embedder in Equation (3.18)), to obtain signals $\mathbf{s_0}$ and $\mathbf{s_1}$. The estimated message bit is then determined by:

$$\hat{m} = \underbrace{argmin}_{b\in 0,1} (r - s_b)^2 \tag{3.27}$$

Once again, we could embed one message bit into a sequence of N samples, a rate $1/N$ repetition encoding can be exploited. Similar to discussions in section 3.1.3, there are two options to do detection: hard or soft decision. The hard decision detector is given by:

$$\begin{aligned} &\hat{m}_n = \lfloor \tfrac{2}{N+1} \textstyle\sum_{h=(n-1)N+1}^{nN} \underbrace{argmin}_{b\in 0,1} (r_h - s_b(h))^2 \rfloor, \\ &n = 1, 2, \ldots, L/N, \end{aligned} \tag{3.28}$$

where L is the length of the signal vector and $\lfloor\ \rfloor$ is the floor function

Alternatively, with the same code rate, the soft decision detector is:

$$\begin{aligned} &\hat{m}_n = \underbrace{argmin}_{b\in 0,1} \textstyle\sum_{h=(n-1)N+1}^{nN} (r_h - s_b(h))^2, \\ &n = 1, 2, \ldots, L/N. \end{aligned} \tag{3.29}$$

In order to ensure the optimality of the system using α_{opt} discussed earlier in this section, the detector must also have knowledge of α_{opt} used in the embedder.

Eggers also proposed a quantization based watermarking scheme called *Scalar Costa Scheme* (SCS) [EBTG03] SCS is a distortion compensated quantization taking advantage of dither signal, which is actually very similar to distortion compensated

dither modulation (DC-DM).

3.3.2 DC-QIM experimental results

To examine the performance of DC-QIM, we performed experiments to study the robustness against additive white Gaussian noise. For the experiments, we took a database of 1,000 images from the Corel database with dimension of 768×512. Thus, there are 6,144 8×8 blocks for each of the image. A binary message of length $12,288$ bits is embedded into each image, which is equivalent to embedding 2 bits per block. We extracted 62 DCT coefficients from each 8×8 block, ignoring the DC and highest frequency coefficients. The entire sequence of 62×6144 coefficients were then pseudo randomized and each bit of the message was embedded in 31 random coefficients. This means that $N = 31$ in Equations (3.28) and (3.29)

We fixed embedding step size to be 3. Consequently, the expected embedding-induced distortion is $D_e = \Delta^2/12 = 0.75$, independent of compensation parameter α according to Equation (3.23).

Under the experimental conditions described above, we set α to a variety of values, and observed that the embedding-induced distortion remains the same. In practice, the DWR was measured at $35dB$ and the measured $D_e = 0.79$, which is slightly higher than the theoretical value $D_e = 0.75$. This is due to the fact that our experiments are performed in the DCT domain. To be transfered back to pixel domain, the DCT coefficients of quantization outputs undergo Inverse DCT and then need to be rounded to integer pixel values. These truncation errors cause the difference between theoretical and practical value of D_e.

The purpose of the experiment was to verify that the optimal compensation factor actually performs best in terms of received bit-error-rate (BER). The distortion compensation factor α was calculated assuming additive white Gaussian noise with a standard deviation of 1.0. In this case, assuming $\mathbf{y}$ is the watermarked signal output from embedder, as in Equation (3.18), the detector received the signal $\mathbf{r} = \mathbf{y} + \mathbf{v}$, where $\mathbf{v}$ is normal distributed noise with standard deviation $\sigma_v = 1.0$. Under this condition, the

optimal α, denoted $\alpha_{opt|\sigma_v=1.0}$, can be calculated using Equation (3.25) to give:

$$\begin{aligned} \alpha_{opt|\sigma_v=1.0} &= \frac{D_e}{D_e + \sigma_v^2} \\ &= \frac{0.79}{0.79 + 1} \\ &= 0.443 \end{aligned} \tag{3.30}$$

We can also calculate the optimal compensation factor for additive white Gaussian noise with the standard deviation of 1.3, $\alpha_{opt|\sigma_v=1.3}$.

$$\begin{aligned} \alpha_{opt|\sigma_v=1.3} &= \frac{D_e}{D_e + \sigma_v^2} \\ &= \frac{0.79}{0.79 + 1.69} \\ &= 0.32 \end{aligned} \tag{3.31}$$

Experiments of robustness versus additive white Gaussian noise were done using these two values of α_{opt} for DC-QIM. For comparison, curves of $\alpha = 0.30$, 0.40 and 0.50 are also examined. Results are shown in Figure 3.2.

Figure 3.2 shows that BER curves cross one another as a function of AWGN. It is observed that given $\sigma_v = 1.0$, the curve of $\alpha = 0.443$ has the lowest BER and given $\sigma_v = 1.3$, the curve of $\alpha = 0.32$ has the best performance.

3.4 Spread Transform Dither Modulation

Quantization index modulation (QIM) is a popular form of digital watermarking based on the framework of communications with side information [Cos83]. In their original paper, Chen and Wornell [BG01a] described a number of variants of the basic QIM algorithm, namely dither modulation QIM (DM), distortion-compensated QIM (DC-QIM), which can find optimal balance point between robustness and capacity under AWGN attack. Furthermore, the distortion compensated version can also be com-

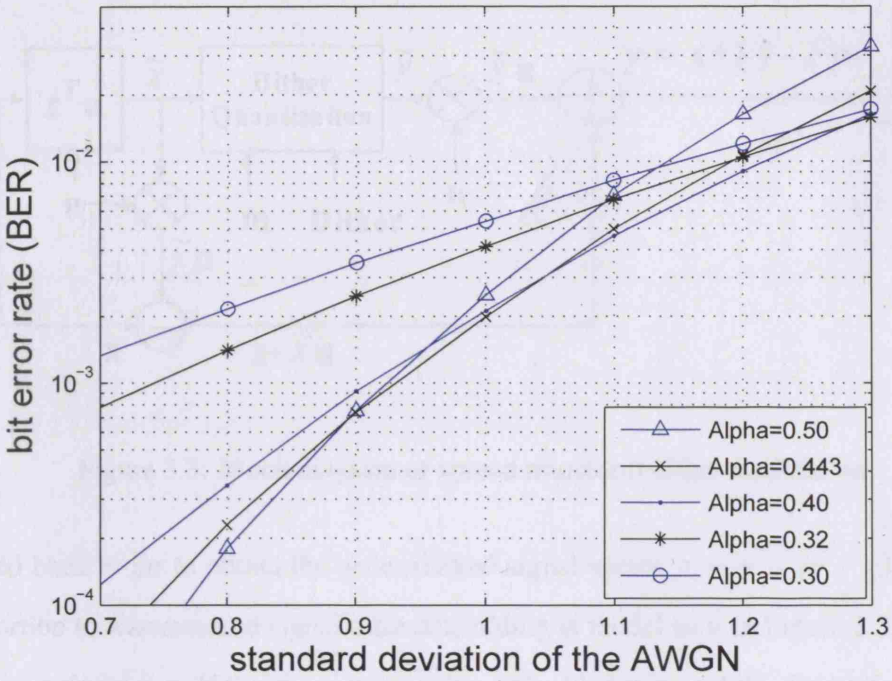


Figure 3.2: Bit error rate(BER) as a function of additive white Gaussian noise for various parameters

bined with DM, yielding distortion compensated dither modulation (DC-DM). Alternatively, the correlation between the host signal sequence and a reference pseudo-noise sequence is quantized [BG01a, PGBH03], leading to the spread transform dither modulation (STDM) scheme that is discussed in this section.

Figure 3.3 illustrates the basic framework for spread transform dither modulation.

STDM differs from regular QIM in that before any quantization operation, the original signal vector, $\mathbf{x}$, is first projected onto a randomly generated vector, $\mathbf{u}$. In Figure 3.3, the *dot product* between vector $\mathbf{x}$ and vector $\mathbf{u}$ is computed as $\tilde{x} = \mathbf{x} \cdot \mathbf{u} = \mathbf{x}^{\mathrm{T}}\mathbf{u}$. Assuming $\mathbf{u}$ is a *unit vector*, the vector $\tilde{x}\mathbf{u}$ is consequently the projection vector of the vector $\mathbf{x}$ onto the vector $\mathbf{u}$ and the scalar $\tilde{x}$ is the length of this projection. As shown in Figure 3.3, this resulting scalar value $\tilde{x}$ is inputted into the dither quantization box. The dither quantization output $\tilde{y}$ is then used to generate the vector $\tilde{y}\mathbf{u}$ which can be considered a replacement of the projection vector $\tilde{x}\mathbf{u}$ in order to carry the message m. Note that the components of vector $\mathbf{x}$ orthogonal to vector $\mathbf{u}$ is $\mathbf{x} - \tilde{x}\mathbf{u}$ which are

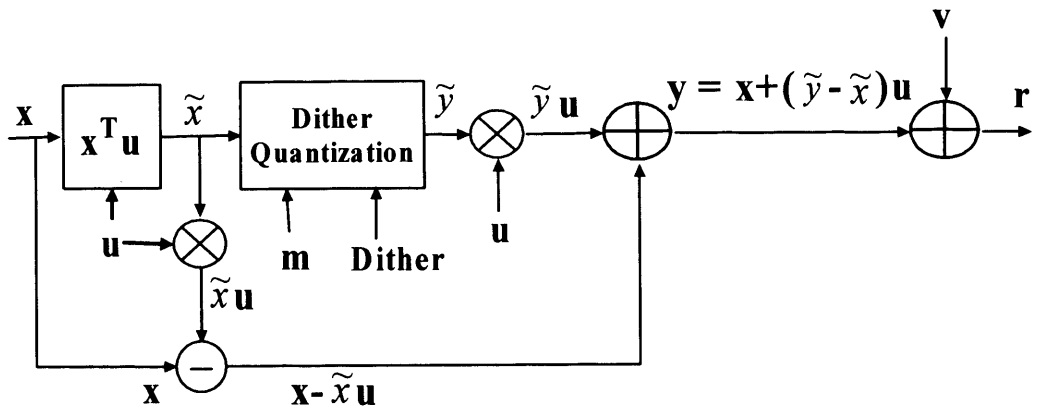


Figure 3.3: Block diagram of spread transform dither modulation

added back to $\tilde{y}\mathbf{u}$ to obtain the watermarked signal vector $\mathbf{y} = \mathbf{x} + (\tilde{y} - \tilde{x})\mathbf{uo}$. The distortion to watermarked signal after embedding is model as $\mathbf{v}$ in Figure 3.3, and the received signal is $\mathbf{r}$. If the message bit to be embedded is m and the dither signal used for dither quantization is d, the equation for STDM embedding is then given by:

$$\mathbf{y} = \mathbf{x} + \left(Q(\mathbf{x}^{\mathrm{T}}\mathbf{u}, m, \Delta, d) - \mathbf{x}^{\mathrm{T}}\mathbf{u}\right)\mathbf{u}, \quad m \in \{0, 1\} \tag{3.32}$$

and the corresponding detection is:

$$\hat{m} = \arg \min_{b \in \{0,1\}} \left(\mathbf{r}^{\mathrm{T}}\mathbf{u} - Q(\mathbf{r}^{\mathrm{T}}\mathbf{u}, b, \Delta, d)\right)^2 \tag{3.33}$$

where b is an assuming bit during detection.

3.5 Problems of QIM

The popularity of QIM is, in part, due to its ease of implementation, computational efficiency and amenability to theoretical analysis. Nevertheless, there are practical limitations of QIM. For example, QIM may introduces perceptually visible distortion due to the quantization step size used. Even worse, QIM is extremely sensitive to valumetric scaling and re-quantization. Valumetric scaling is a very common signal processing operation and occurs whenever the volume of an audio signal or the brightness of an

image is changed. Re-quantization is also common and occurs when any multimedia digital signal undergoes digital-to-analog conversion and subsequent analog-to-digital conversion. A major application of watermarking is to provide protection from this "analog hole". Thus, if QIM methods are to be used it is imperative that they be robust to re-quantization. Even in the absence of D-to-A and A-to-D conversion, re-quantization will occur whenever a multimedia signal undergoes lossy compression or numerical rounding.

3.5.1 Vulnerability to amplitude scaling

In theoretical discussions of watermarking systems, it is tempting to deal with only additive noise, because this form of distortion is easy to analyze. However, in reality, many, if not most, processes applied to watermarked signal are not well modeled by additive noise. In fact, many signal processings are functions of the content itself. A simple, but important, example is that of amplitude modification. That is,

$$r = \beta y, \tag{3.34}$$

where $\mathbf{y}$ is watermarked signal and β is a scaling factor. As to image and video, this simply means adjusting the brightness and contrast. For audio, represents a change in volume. Obviously, these are very commonly used processes and digital watermarking is expected to survive amplitude modification. However, as stated earlier, QIM algorithms are very sensitive to amplitude scaling.

In order to illustrate the problem we investigated the robustness of dither modulation (DM) to amplitude scaling.

We took a database of 1,000 images, each of dimension 768×512. A binary message of length $12,288$ bits is embedded into each image. We extract 62 DCT coefficients from each 8×8 block. The entire sequence of 62×6144 coefficients are then pseudo-randomized and each bit of the message is embedded in 31 random coefficients. This is equivalent to embedding two bits in each block of the image and thus we consider the embedding rate as $1/32$.

The embedding is performed using the regular DM scheme discussed in this chapter. The quantization step size is fixed to 3.0 and the corresponding DWR is $35dB$ for the watermarked images. Figure 3.4 shows robustness result against amplitude scaling. It is seen that if we adjust the image to be 20% brighter or darker, the bit error rate (BER) is about 10%, and when the scaling factor is 0.5 or 1.5, BER can be as high as 30%.

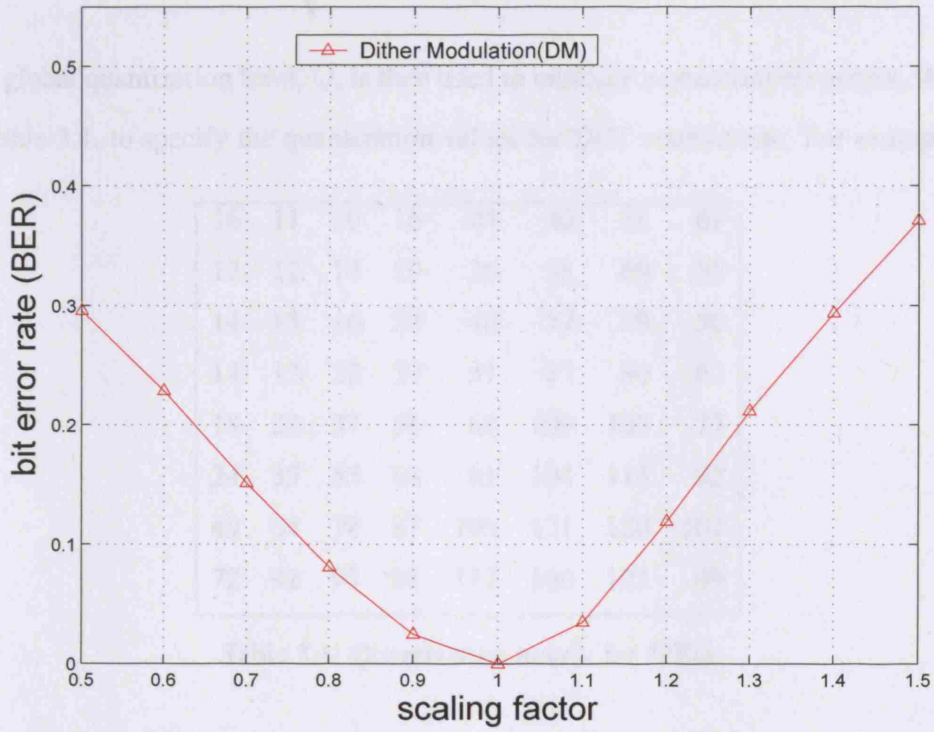


Figure 3.4: Bit error rate(BER) versus amplitude scaling for dither modulation with embedding rate of 1/32 and fixed DWR of 35dB

The reason for this problem is that the quantization step size for DM (or regular QIM) is fixed and does not scale accordingly to the amplitude scaling factor. This motivated us to propose a method with adaptive step size to overcome this limitation of QIM. This method is further detailed in Chapters 4 and 5.

3.5.2 Sensitivity to re-quantization

JPEG compression is a widely used re-quantization processing. We now test the effects of JPEG compression on dither modulation. Different JPEG encoders may have differ-

ent effects since the JPEG standard does not actually specify a compression method. However, it is described in [MDC04] that a conventional method of JPEG encoding is very common. In this method, given a user-specified JPEG *quality factor*, QF, in the range of 0 to 100, a global quantization level, Q, is computed by:

$$Q = \begin{cases} \frac{50}{QF}, & \text{if } QF < 50, \\ 2 - 0.02 \cdot QF, & \text{if } QF \geq 50. \end{cases} \tag{3.35}$$

The global quantization level, Q, is then used to multiply a quantization matrix, shown in Table 3.1, to specify the quantization values for DCT coefficients. For example, if

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Table 3.1: Quantization matrix for JPEG

$QF = 90$, then $Q = 0.2$, the DC term is quantized with a quantization step size of $q = 0.2 \times 16 = 8$ and the highest-frequency term with $q = 0.2 \times 99 = 19.8$.

The effects of JPEG compression can be simulated by applying quantization with the step size of q in the DCT domain. The DCT coefficients after JPEG compression are obtained by the following quantization:

$$c_q = q \cdot round\left(\frac{c}{q}\right) \tag{3.36}$$

where c and c_q are the DCT coefficient before and after JPEG. After quantization, the inverse block-DCT was applied.

It is sufficient to apply this operation to test the effect of JPEG. This is preferable

to using some encoder from a third party, as they may change over time and we cannot be sure of the algorithms used

Both QIM and DM are very sensitive to re-quantization. Figure 3.5 illustrates this point for DM. Here, robustness to JPEG compression is examined for DM in the discrete cosine transform (DCT) domain, i.e. we quantize the DCT coefficients rather than the pixel values. Similar performance was observed in the pixel domain.

For this experiment, the embedding conditions and parameters are same as for the experiment done for amplitude scaling. Testing was performed on 1000 images. If we arbitrarily consider when the BER exceeds 20%, this occurs at a JPEG quality factor of about 94%. These quality factors are extremely high and in many scenarios much better robustness may be needed.

In Chapter 4, we propose a method combining STDM with a perceptual model to address this problem.

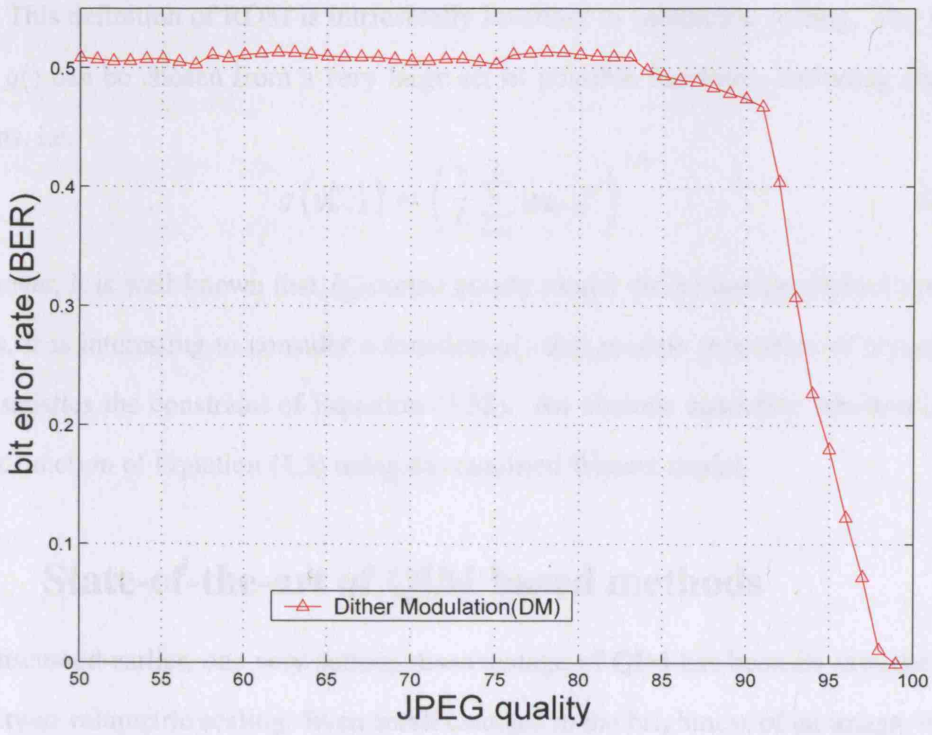


Figure 3.5: Bit error rate(BER) as a function of JPEG quality for dither modulation with embedding rate of 1/32 and fixed DWR of 35dB

3.6 Rational Dither Modulation

Rational dither modulation (RDM) was first proposed by Perez-Gonzalez *et al.* [PG-BAM04] and is intended to provide valumetric invariance to QIM. Given a host signal, $x = (x_1 \dots x_N)$ and a watermarked signal, $y = (y_1 \dots y_N)$, then the k-th bit of a watermark message, $m = (m_1 \dots m_M)$, is embedded as

$$y_k = g\left(y_{k-L}^{k-1}\right) Q_{m_k}\left(\frac{x_k}{g(y_{k-L}^{k-1})}\right) \tag{3.37}$$

where y_{k-L}^{k-1} denotes the set of past watermarked signals, $(y_{k-L} \dots y_{k-1})$ and the function, $g()$ maps its L-dimensional input vector to a real value and has the property that for any valumetric scaling factor $\beta > 0$,

$$g(\beta y) = \beta g(y) \tag{3.38}$$

This definition of RDM is intrinsically invariant to valumetric scaling. The function $g()$ can be chosen from a very large set of possible functions, including the L_p-norms, i.e.

$$g\left(y_{k-L}^{k-1}\right) = \left(\frac{1}{L}\sum_{i=1}^{L}|y_{k-i}|^p\right)^{1/p} \tag{3.39}$$

However, it is well-known that L_p-norms poorly model the human perceptual system. Thus, it is interesting to consider a function $g()$ that models properties of perception and satisfies the constraint of Equation (3.38). An obvious candidate function is the slack function of Equation (4.3) using the modified Watson model.

3.7 State-of-the-art of QIM based methods

As discussed earlier, one very serious disadvantage of QIM has been its extreme sensitivity to valumetric scaling. Even small changes in the brightness of an image, or the volume of a song, can result in dramatic increases in the bit error rate. This valumetric scaling is a very common occurrence and there has therefore been considerable work addressing this issue [EBG02, LKKM03, OKS04, Bas05, LS04, SLH04].

Eggers *et al* [EBG02] proposed to estimate the valumetric scaling by "securely embedd[ing] SCS pilot watermark". However, the fact that all watermarked content may contain the same pilot signal may lead to a security weakness – if the pilot signal can be estimated and removed, the watermark may not be detected. At the very least, the pilot signal is likely to reduce the watermark payload. Lee *et al.* [LKKM03] proposed estimating the global scaling factor using an EM algorithm, which does not need a pilot watermark. However, they note that the "complexity could be impractical". The closest work to ours is that of Oostveen *et al* [OKS04] which uses a simple perceptual model based on Weber's law. The quantization step size is a function of the average brightness of a neighborhood of pixels. This provides a simple perceptual model in which bright regions undergo larger changes than dark regions. It is obvious that if the image is scaled by a factor, β, then the quantization step size is scaled similarly.

Lagendijk *et al.* [SLH04,LS04] have presented several methods based on the characteristic function of the signal, and on a maximum likelihood procedure. Their algorithm requires models of both the host signal and noise. Experimental results are reported on synthetic data and real audio data. However, it is unclear how the algorithms will perform on real imagery.

Bas [Bas05] recently proposed a method using so called "floating quantizers". The quantization, step sizes are based on the minimum and maximum of a triplet of pixels. A key advantage of this method is its robustness to non-linear valumetric scaling such as gamma correction.

QIM based approaches have practical limitations due to its extreme sensitivity to valumetric scaling and re-quantization. The preceding chapter mainly concentrated on improving the robustness to valumetric scaling, which is a very common signal processing operation and occurs whenever the volume of an audio signal or the brightness of an image is changed. Re-quantization is also common and occurs when any multimedia digital signal undergoes digital-to-analog conversion and subsequent analog-to-digital conversion. A major application of watermarking is to provide protection from this "analog hole". Thus, if a watermarking scheme is to be used for this application it is

imperative that it be robust to re-quantization. Even in the absence of D-to-A and A-to-D conversion, re-quantization will occur whenever a multimedia signal undergoes lossy compression or numerical rounding.

The problem of valumetric scaling has received widespread attention and a number of solutions have been proposed [PGMBA05, LC05b, LC05a, SL06]. In contrast, there has been surprisingly little research focused on the issue of re-quantization, among which JPEG compression is a typical one.

Fei *et al.* [FKK04] analyzed the performance of two popular classes of watermark embedding techniques, spread spectrum watermarking and quantization-based embedding, in the presence of JPEG compression. They also proposed a hybrid watermarking scheme to exploit the theoretically predicted advantages of spread spectrum and quantization-based watermarking to achieve superior performance. In contrast, we focus on improving the fidelity and/or robustness of STDM to both re-quantization and valumetric scaling in this chapter.

Pérez-Gonzàlez *et al.* [PGCB03] examined the performance of Distortion Compensated Dither Modulation (DC-DM) against JPEG compression and proposed a new method for detection based on a weighted Euclidean distance. Experimental results demonstrated improved performance over traditional DC-DM. However, there is no comparison with STDM and it remains unclear whether this method is superior to STDM.

Chapter 4

Improving QIM methods by Using a Modified Perceptual Model

Perceptual models are very important to digital watermarking, not only because they can measure the quality and differences between multimedia signals, but also because they can help improve watermarking designs.

Each watermarking application may have its own specific requirements but the two most important and mutually conflicting requirements are usually *fidelity* and *robustness*. It is well-known that improvements in fidelity and robustness of watermarking schemes can be achieved by adapting the watermark strength to the local perceptual characteristics of the digital contents. To address the fidelity issue, in this chapter we apply Watson's perceptual model to determine by how much each DCT coefficient can be altered. This quantity, referred to as "slack", is then used to adaptively adjust the quantization step sizes used to quantize the DCT coefficients. This adaptive algorithm is described in Section 4.1. This algorithm provides significant improvements in fidelity, as measured by Watson distance, and performance degrades more gracefully with additive white Gaussian noise. However, this adaptive method remains vulnerable to amplitude scaling as traditional DM does.

The slacks computed by Watson's model do not scale linearly with valumetric scaling. As a result, our adaptive DM method proposed in Section 4.1 is not robust to valumetric scaling. In order to be robust to valumetric scaling it is necessary that

the quantization step sizes be scaled by the same scaling factor that the signal has undergone in order to correctly perform QIM decoding. In this chapter, we propose a small modification to the Watson model and this modification leads to slacks that do scale linearly. This algorithm, referred to as the dither modulation based on modified Watson (DM-MW) is compared to the work of Oostveen *et al* [OKS04] and shown to have superior performance. The modification of the Watson model does lead to a degradation in fidelity. However, this degradation is slight and the new algorithm still has significantly better fidelity when compared with either the original DM algorithm or that of Oostveen *et al* [OKS04]. This work is described in Section 4.3. To further reduce the source of errors, we have also combined RDM with Modified Watson, showed in Section 4.4.

A new method, referred to as STDM-W, is then discussed in Section 4.5 to describe how the projection vector used in STDM can be chosen so as to minimize the perceptual distortion. However, the STDM-W algorithm remains sensitive to valumetric scaling attack. In Section 4.6, we first propose an adaptive STDM method (STDM-MW-SS) to overcome this problem. STDM-MW-SS uses a modified perceptual model that scales linearly with amplitude (valumetric) scaling not only to select the projection vector but also to calculate the quantization step size. This improves both the fidelity (compared to STDM) and provides robustness to valumetric scaling. We then demonstrate that such an approach can be extended to another method, STDM-OptiMW-SS, which is able to offer significant improvements in robustness to JPEG compression.

4.1 Adaptive Dither Modulation Based on Watson's Perceptual Model

In previous Sections 2.8, 2.9 and 2.10, several perceptual models were discussed. It is widely acknowledged that the ability to perceive a change depends on the content. For example, the human visual system is much less sensitive to changes in heavily textured image regions and much more sensitive to changes in uniform regions. On the

other hand, a watermarking embedder introduces changes that have an impact on the fidelity of images. It is well recognized that fidelity and robustness of watermarking can be improved by adapting the watermark strength (the changes) to the local perceptual characteristics of the digital contents. However, the original QIM and DM algorithms, described in Chapter 3, use a fixed quantization step size that is independent of the content. To account for this, we propose using the Watson's model, discussed in Section 2.10, to automatically select the quantization step size at each sample.

Watson's model is chosen for our research because: (i) It has been previously used and well recognized by the watermarking community (ii) It is easy to implement in DCT, this is useful as many compression algorithms work in the DCT domain (iii) Thesis shows that Watson's model can be easily modified to scale linearly with amplitude scaling, which solves one of QIM's problem and is discussed further in Section 4.3

In this section, we are able to provide significantly improved fidelity, by locally adapting the quantization step size. For example, in regions of high texture, a larger step size can be used, while in regions of low texture, a small step size is chosen. As we shall demonstrate, this adaptivity can simultaneously improve robustness, at least in high noise regimes.

Dither modulation (DM) is robust to additive white Gaussian noise, provided the standard deviation of the noise remains small compared with the quantization step size. For dither modulation (discussed in Section 3.2), the distance between $d(n,0)$ and $d(n,1)$ (defined in Equation 3.13) is $\Delta/2$, i.e. the distance between the embedding output for message bit '0' and '1' is $\Delta/2$. Thus, if the noise exceeds $\Delta/4$, the bit error rate (BER) for DM rapidly degrades.

In contrast, an adaptive step size has, by definition, many different step sizes. Thus, we would expect the rate of change in BER to be slower. This is experimentally confirmed.

Traditional DM is *non-adaptive*, using a uniform scalar (step size Δ) for quantization as shown in Equation (3.12). Note that the *slacks* of Watson's model, defined

by Equation (2.23), evaluate the amounts by which individual DCT coefficients may be changed according to Watson's perceptual model. This motivates us to design an adaptive DM method in which the DCT coefficients are quantized using step sizes that are based on Watson's perceptual model. We can use the slacks of Equation (2.23) to adaptively select the quantization step size.

The adaptive DM system is schematically shown in Figure 4.1.

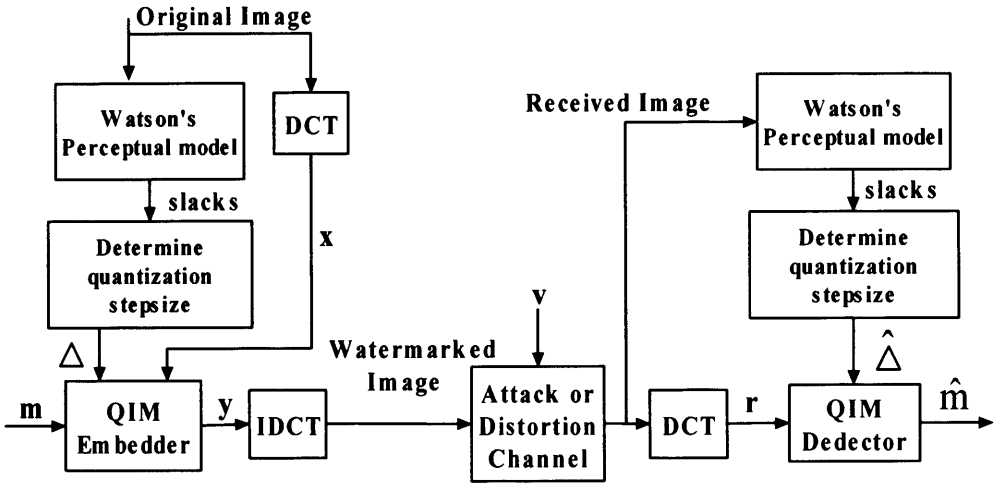


Figure 4.1: Adaptive dither modulation based on Watson's model.

The cover Work is converted to the DCT domain and the coefficients serve as the host signal x. The slacks from Watson model are multiplied by a global constant, G, to determine the final quantization step size Δ for each DCT coefficient. The global constant, G, must be known to the detector and is the equivalent of the detector knowing the fixed quantization step size in traditional DM. The constant, G, is empirically adjusted to control the watermark strength and the document-to-watermark ratio. The message m is embedded by the DM embedder to obtain the watermarked signal y. After transmission, the received Work, r is used to estimate the corresponding quantization steps, $\hat{\Delta}$. The estimation procedure is exactly the same as at the embedder. However, the estimate is now applied to the received, watermarked Work rather than the original, unwatermarked Work. Finally, using these step sizes, the message $\hat{m}$ is detected by the DM detector using Equations (3.14) and (3.17).

Note that we use the original Work to compute the quantization step size for each sample during embedding, and we use the distorted watermarked Work to compute the quantization step size for each sample during detection. If these two step sizes are not the same, then a bit error may occur. In fact, even without distortions, there is the possibility that the slacks, and therefore the quantization step sizes, computed at the detector will be different, due to the changes introduced by the embedded. However, in practice, very good correspondence is achieved, as is demonstrated below.

4.2 Vulnerability of the Adaptive Scheme to Amplitude Scaling

Unfortunately, despite this new adaptive method's superior performance under additive white Gaussian noise conditions (as illustrated in Section 5.2), it remains vulnerable to *amplitude scaling*. When the amplitude of image is scaled by factor of β, the resulting luminance-masked threshold (denoted as $\widehat{t_L}[i,j,k]$) is calculated as:

$$\widehat{t_L}[i,j,k] = t[i,j](\frac{\beta C_o[i,j,k]}{\beta C_{0,0}})^{\alpha_T} = t_L \tag{4.1}$$

that is, $\widehat{t_L}$ does not scale linearly with amplitude scaling, but is, in fact, invariant to amplitude scaling. Thus, referring to Equation (2.23), the slack and corresponding quantization step size, $\hat{\Delta}$, are not proportional to scaling factor β, and the adaptive DM method is therefore sensitive to amplitude scaling.

We have examined the robustness of traditional dither modulation (DM) to amplitude scaling in Section 3.5.1. For comparison, we now investigate our adaptive DM. The experimental conditions and parameters are same as for the experiment done in Section 3.5.1 and the testing was performed on 1000 images. The robustness to amplitude scaling for both schemes is shown in Figures 4.2.

From Figures 4.2, we observe that these two schemes perform similarly in terms of bit error rate(BER) versus amplitude scaling. When $0.9 \leq \beta \leq 1.1$, the original DM performs as well or slightly better than our adaptive scheme, but for larger scale

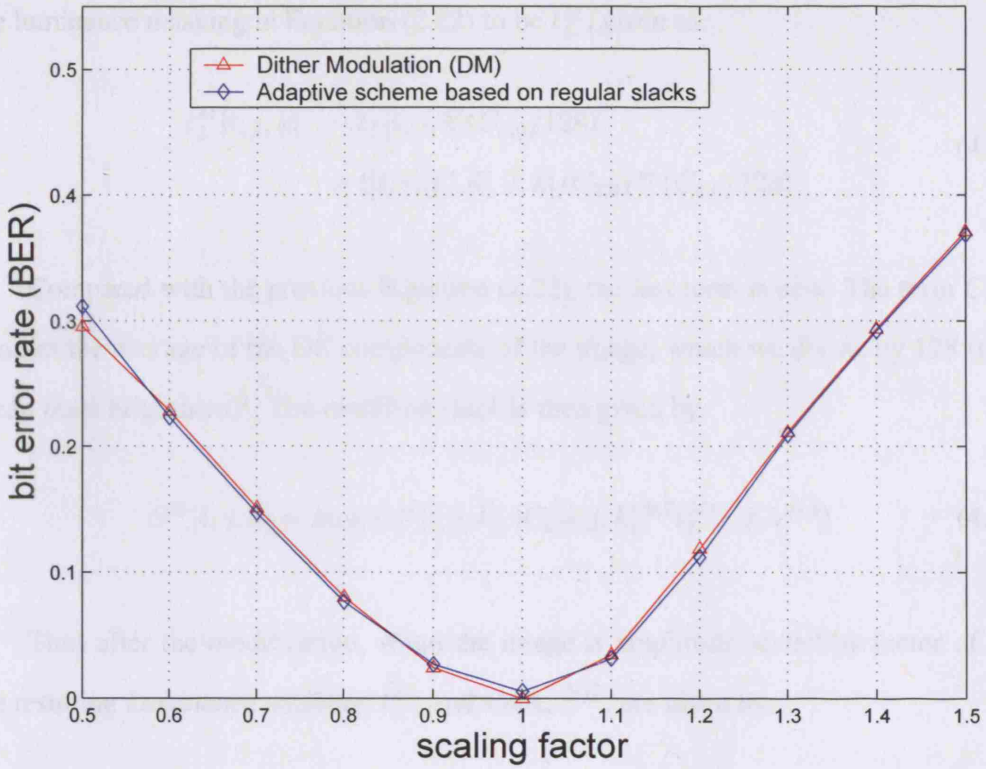


Figure 4.2: Bit error rate(BER) versus amplitude scaling with embedding rate of 1/32 and fixed DWR of 35dB

this adaptive method has similar or superior performance. However, if we adjust the image to be 20% brighter or darker, the BER for both methods are about 10%. When the scaling factor is 0.5 or 1.5, the BERs can be as high as 30%.

4.3 Adaptive Dither Modulation Based on a Modified Watson Model

The previously described adaptive DM algorithm based on Watson's perceptual model (proposed in Section 4.1) remains sensitive to valumetric scaling, since the slacks do not scale linearly with amplitude scaling. To be robust to valumetric scaling, we need the slacks to scale linearly with valumetric scaling, i.e. we want the estimated $\hat{\Delta}$ to be multiplied by β when the amplitude of the signal is scaled by β. To this end, we modify

the luminance masking in Equation (2.22) to be t_L^M, given as:

$$\begin{aligned} t_L^M[i,j,k] &= t_L[i,j,k](C_{0,0}/128) \\ &= t[i,j](C_o[0,0,k]/C_{0,0})^{\alpha_T}(C_{0,0}/128) \end{aligned} \tag{4.2}$$

Compared with the previous Equation (2.22), the last term is new. The term $C_{0,0}$ denotes the average of the DC components of the image, which we divide by 128 (the mean pixel brightness)[1]. The modified *slack* is then given by:

$$S^M[i,j,k] = \max\left(t_L^M[i,j,k], |C_o[i,j,k]|^{0.7} t_L^M[i,j,k]^{0.3}\right) \tag{4.3}$$

Thus after the modification, when the image is amplitude scaled by factor of β, the resulting *Luminance masking*, $\widehat{t_L^M}$, and *slack*, $\widehat{S^M}$, are given by:

$$\begin{aligned} \widehat{t_L^M}[i,j,k] &= t_L[i,j,k](\beta C_{0,0}/128) \\ &= t[i,j](\beta C[0,0,k]/\beta C_{0,0})^{\alpha_T}(\beta C_{0,0}/128) \\ &= \beta t_L^M[i,j,k] \end{aligned}$$

$$\begin{aligned} &\widehat{S^M}[i,j,k] \\ &= \max\left(\widehat{t_L^M}[i,j,k], |(\beta C_o[i,j,k])|^{0.7} \widehat{t_L^M}[i,j,k]^{0.3}\right) \\ &= \max\left(\beta t_L^M[i,j,k], \beta^{0.7}|C_o[i,j,k]|^{0.7}\beta^{0.3} t_L^M[i,j,k]^{0.3}\right) \\ &= \beta \max\left(t_L^M[i,j,k], |C_o[i,j,k]|^{0.7} t_L^M[i,j,k]^{0.3}\right) \\ &= \beta S^M[i,j,k] \end{aligned} \tag{4.4}$$

In the modified Watson model the new luminance masking t_L^M and slack S^M scale linearly with β. The modified slack can then be used to determine the step size Δ_n^M.

$$\Delta_n^M = G \times S_n^M, n = 1, 2, \ldots, L. \tag{4.5}$$

[1]Note that since the Watson distance and the proposed modified Watson distance are a function of $C_{0,0}$, the average brightness of the image, our methods will be susceptible to any cropping operation that changes the overall brightness value.

When the image is scaled by factor of β, the estimated quantization step size $\widehat{\Delta_n^M}$ is also scaled by β. This provides an adaptive QIM algorithm that is theoretically invariant to valumetric scaling.

4.4 RDM with Modified Watson's Model

Dither modulation using a Modified Watson model (DM-MW), discussed in Section 4.3, implicitly assumes that the slacks calculated by the embedder and detector are the same, even though the function of embedding alters the DCT coefficients. Since these alterations are small, this assumption is usually true. However, this is also the source of some error. To guarantee that the slack is unaffected by the embedding procedure, we investigated using rational dither modulation in current Section.

In Section 4.3, we used the modified slack to adaptively set the quantization step size and thereby provide robustness to valumetric scaling. Figure 4.1 is a block diagram of the system.

Rational dither modulation (RDM) [PGBAM04] has been recently proposed as an alternative DM method in which the quantization step size at time, k, is a function of the *watermarked* samples at earlier times. This algorithm is described in Section 3.6. If this function is chosen such that it scales linearly with amplitude, then RDM is invariant to valumetric scaling. To incorporate a perceptual model within the RDM framework, we propose to calculate the quantization step sizes for the current block, k, based on slacks of the previous watermarked block. Thus, while the slacks of the current block k are affected by the embedding process itself, the embedding of block k is based on the slacks from the previously watermarked block $k-1$, whose slacks were altered in the previous iteration but are unaffected by the processing of block k. Thus, both the watermark embedder and the watermark detector are guaranteed to use the identical values of step size (ignoring noise and round-off error). This is described in Section 4.4. Of course, once again there is a degradation in fidelity, but experimental results indicate that it is small.

Notice that the quantization step size is determined by a local neighborhood

around the host sample, x_k, and that this neighborhood is altered during the embedding of the watermark. Thus, during detection, we must rely on the fact that these alterations are small, and hope that the slacks based on the modified local neighborhood are the same as those determined during embedding. While this is often true, rational dither modulation suggests an alternative approach, in which the perceptual slack at time k is based on a nearby neighborhood of previously watermarked samples. Clearly, there may be some degradation in perceptual quality since a perceptual estimate made in a nearby neighborhood, is not guaranteed to be perceptually relevant.

4.4.1 Implementation of RDM-MW

Before describing the implementation of rational dither modulation with a modified Watson perceptual model, we first describe an implementation of RDM in the DCT domain. This algorithm, denoted RDM is used for comparison.

Our implementation of RDM quantizes the 62 DCT coefficients of each 8×8 block (excluding the DC and highest frequency terms). For each DCT coefficient, we use its corresponding DCT coefficient from the previously watermarked 8×8 neighboring block to determine the quantization step size. Thus, the window size is 1. The function $g()$ is chosen to be the absolute value of the DCT coefficient, i.e. we use an L_1-norm in Equation (3.39), scaled by a global constant that is chosen so that the document-to-watermark ratio (DWR) averaged over all watermarked images equals a desired value.

We believe that a window size of 1 provides the fairest comparison with our adaptive RDM algorithm. However, we also implemented an RDM algorithm with a window size of 62, denoted RDM-62-L2-Norm, that uses an L2-norm of the 62 DCT coefficients in the previous block.

Our perceptually adaptive RDM method is denoted RDM-MW. To incorporate a perceptual model within the RDM framework, we propose to calculate the quantization step sizes for the current block, k, based on the slacks of the previous watermarked block. That is, each DCT coefficient in the current block is quantized based on the slack of its corresponding coefficient in the previously watermarked block. Thus, this

method also has a window size of 1.

The slacks used to quantize block k are unaffected by the embedding process, since they are determined from the previously watermarked block. Thus, both the watermark embedder and the watermark detector are guaranteed to use the identical values of quantization step size (ignoring noise and round-off error). Of course, we expect a degradation on fidelity since we are basing our perceptual model of block k on calculations performed on block $k-1$. However, provided there is sufficient spatial continuity in the image, then this degradation is likely to be small. This is supported by the experimental results.

For all methods, a message of length 12,288 is embed using a 1/31 rate repetition code, i.e. one message bit is embedded in 31 DCT coefficients. As noted previously, randomization of the DCT coefficients significantly improves performance. However, randomization of the DCT coefficients is problematic for RDM. This is because the adaptive step size depends on the previous block of coefficients. Therefore, the previous neighboring block of coefficients must be watermarked prior to the current block, i.e. not in random order. In [LC05a] we proposed a solution to this problem based on partitioning the image into disjoint regions, selecting a random block from each region, and only randomizing the coefficients in these blocks.

As we embed one bit in 31 DCT coefficients. These coefficients are randomized as described below for the new algorithm, RDM-MW:

1. The DCT coefficients are randomized. In [LC05b] this randomization was over all the coefficients in the image. Here, because we require access the neighboring block in order to computer the slacks, we randomize the DCT coefficients from 32 blocks. See below.

2. Each image is (i) divided into 32 disjoint regions, (ii) a random block in each region is assigned as the start block, (iii) blocks are indexed in a zig zag order in both the positive and negative directions, as depicted in Figure 4.3. The perceptual slack at block k is computed from its neighbouring block $k-1$. The scan

initially precedes in each partition from a random block 1 towards the right. On reaching the end of the partition, the scan then precedes in direction left of the initial block. At any iteration k, the DCT coefficients from the 32 corresponding blocks of the 32 partitions are randomized.

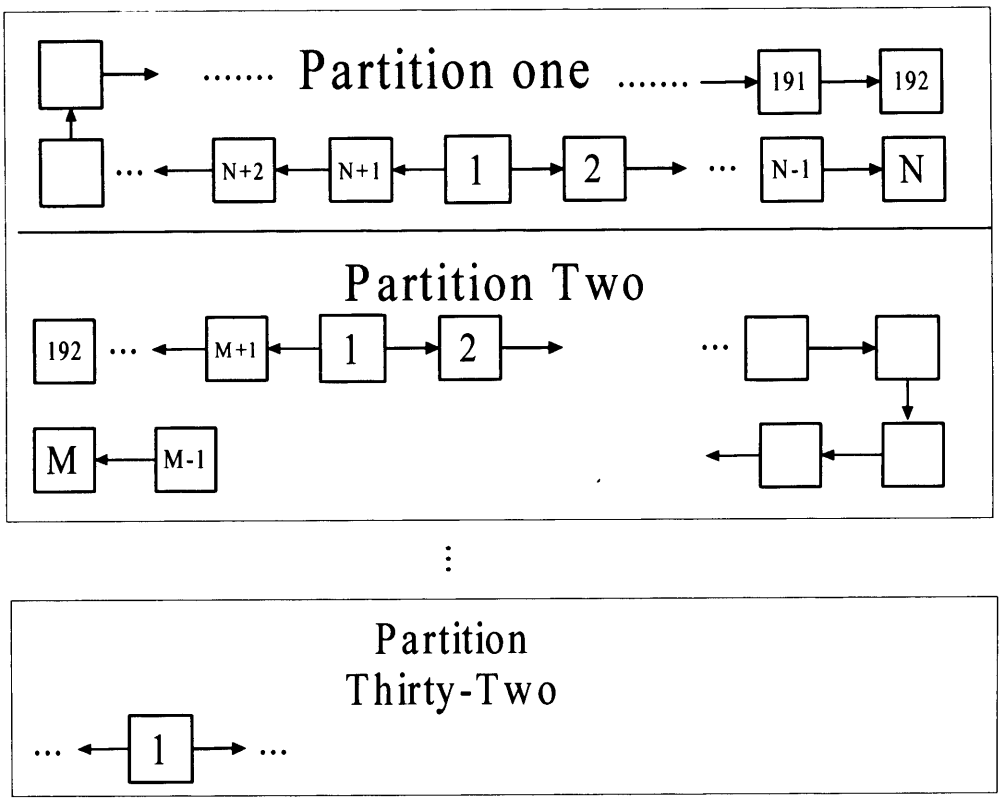


Figure 4.3: Order of scanning blocks for embedding

To embed a bit m_i, 31 random coefficients from the 32 blocks are assigned to the message bit and quantized as in Equation (3.12). The quantization step size for each coefficient is determined by the slack associated with the corresponding coefficient from the previous block. These slacks are multiplied by a global constant that is chosen to provided a desired DWR for the watermarked images.

While very satisfactory results were obtained by the randomization method above, it was pointed out that a simpler solution is to randomize the message code, prior to embedding. Thus, we randomize the 12288×31 length repetition code and sequentially embed it in the DCT coefficients. Note that all experiment results in this chapter are

obtained by randomizing the message codes.

4.5 Combining Spread Transform Dither Modulation with a Perceptual Model

It is discussed in earlier chapters that QIM methods are very sensitive to valumetric distortion. A number of algorithms have been proposed to counter this [PGMBA05, LC05b, LC05a, SL06], specifically rational dither modulation (RDM) [PGMBA05], adaptive QIM using a modified Watson distance (QIM-MW) [LC05b] (Section 4.3) and adaptive RDM using a modified Watson distance (RDM-MW) [LC05a] (Section 4.4). These latter methods are based on adaptively changing the quantization step size, and consequently provide robustness to valumetric scaling. However, experimental results show that these adaptive schemes remain vulnerability to re-quantization, e.g. JPEG compression.

Of the number of QIM variants, spread transform dither modulation (STDM) exhibits most robustness to re-quantization. STDM differs from regular QIM in that the quantization is performed to the projection of the host signal onto a randomly generated vector. Figure 4.4 illustrates this basic idea.

In Figure 4.4, the host signal $\mathbf{x}$ is first projected onto a randomly generated vector, $\mathbf{u}$, to obtain a scalar value, $\tilde{x}$. Assuming the message bit to be embedded is '1' (similar operations apply to bit '0'), the value $\tilde{x}$ is then quantized to the nearest crossing point representing '1' along the axis of $\mathbf{u}$, resulting in a scalar value $\tilde{y}$. The components of the signal $\mathbf{x}$ that are orthogonal to $\mathbf{u}$ is $\mathbf{x} - \tilde{x}\mathbf{u}$. It is added back to $\tilde{y}\mathbf{u}$ to obtain the watermarked signal vector $\mathbf{x}$. Thus the distortion (differences) between $\mathbf{y}$ and $\mathbf{x}$ is $(\tilde{y} - \tilde{x})\mathbf{u}$. The equation for STDM embedding is given by:

Figure fig:STDMvectors shows that the change to the signal $\mathbf{x}$ is in the direction of the random vector $\mathbf{u}$, and the magnitude of the change is controlled by the quantization error. Since $\mathbf{u}$ is random, no consideration is given to the perceptual qualities of the signal $\mathbf{x}$.

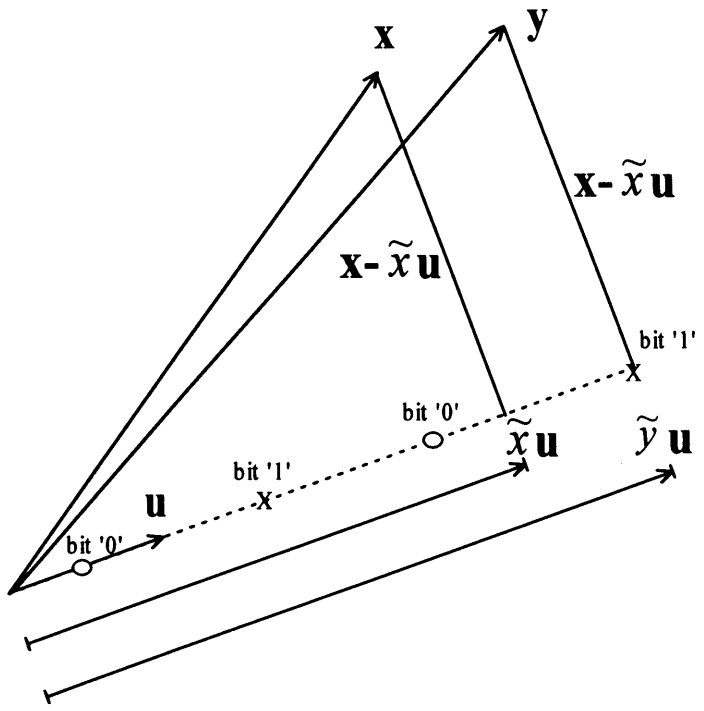


Figure 4.4: Vector operations of spread transform dither modulation

In principle, a perceptual model can provide an estimate of the smallest change that each component of the signal x accepts before becoming just noticeable. In prior work, we have referred to the change needed to introduce a just noticeable distortion (JND) as the "slack".

In practice, for image signals, Watson provides a perceptual model for calculating the slack associated with each discrete cosine transform (DCT) coefficient within an 8×8 block [Wat93a]. Thus, given an image, x, and its block-based DCT coefficients, X, we can apply Watson's model to compute the corresponding slack vector associated with each DCT coefficient. The larger an element of this vector, the more we may change the corresponding DCT coefficient before the change becomes noticeable.

When we look at the original STDM algorithm depicted in Equation (3.32) and Figure 4.4, the embedding algorithm simply reduces to "modify the original Work x so that its projection onto some random vector u is equal to a given value". An obvious and straightforward solution is to move the original Work along the randomly generated

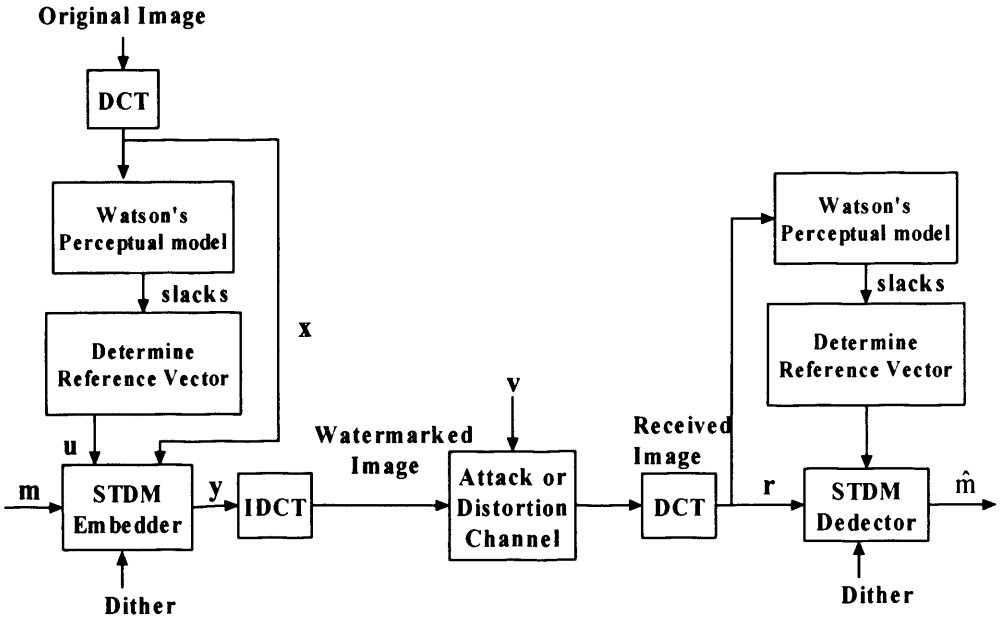


Figure 4.5: Block diagram of STDM watermark embedder and detector with a perceptual model.

projection direction $\mathbf{u}$. However this is not mandatory, the randomly generated $\mathbf{u}$ can be replaced by more meaningful vector. For example, to incorporate a perceptual model within the STDM framework, we propose a new method that the projection vector, $\mathbf{u}$, is assigned the Watson's perceptual slack values corresponding to each DCT coefficient, rather than pseudo-random values. Note that the vector magnitude is normalized to unity and quantization is performed in the DCT domain, as illustrated in Figure 4.5.

In this arrangement, which we refer to as STDM-W (STDM Watson), the change in $\mathbf{x}$ is no longer randomly distributed, but is arranged based on the perceptual properties of the signal - more change is directed to coefficients with larger slacks. As a result, the perceptual distortion introduced by STDM will be substantially reduced, as is confirmed by experiments described shortly.

Since the projection vector is now a function of the signal (image), it is unique for each image. Consequently, a blind watermark detector must be able to estimate the projection vector from the received, watermarked signal, as illustrated in Figure 4.5. However, since watermark embedding alters the signal, the detector's estimate of the

projection vector may not be exact.

4.6 Improving the Robustness to Valumetric Scaling and JPEG Compression

The proposed STDM algorithm based on the Watson's model(STDM-W), while exhibiting improved fidelity, is not invariant to valumetric scaling. This is because the quantization step size for both STDM and STDM-W is fixed, i.e. it does not scale linearly with the valumetric scaling factor, β.

To provide robustness to valumetric scaling, we need to ensure: (i) that the step size scales linearly with valumetric scaling, i.e. we want the estimated $\hat{\Delta}$ to be multiplied by β when the amplitude of the signal is scaled by β, and (ii) the reference vector $\mathbf{u}$ used in embedder (Equation (3.32)), has (approximately) the same direction as $\hat{\mathbf{u}}$ used in detector (Equation (3.33)). Note, however, that $\mathbf{u}$ and $\hat{\mathbf{u}}$ do not have to be identical, though some small degradation in performance may occur.

In our previous work [LC05b] and Section 4.3, we proposed a modified Watson's model such that the modified slack, S^M, scales linearly with β. Based on this perceptual model, we designed three STDM based algorithms. The first, STDM-MW, simply replaces the Watson model of STDM-W with the modified perceptual model and is provided for evaluation purposes. The second, STDM-MW-SS, also adaptively modifies the quantization step size thereby providing invariance to valumetric scaling. A further modification to Watson's model, provides us with a third STDM method, STDM-OptiMW-SS. All these new methods are described in this section.

STDM-MW

This method, STDM-MW, is similar to STDM-W, except that the projection vector is now determined by the modified perceptual model. We provide this in order to examine the perceptual impact of the modification.

STDM-MW-SS

The STDM-W and STDM-MW methods do not provide invariance to valumetric scaling. The STDM-MW-SS not only uses the perceptual model to determine the projection vector, but also uses the same model to select the quantization step size.

Given a $length - L$ vector of DCT coefficients $\{x_i; i = 1, 2, \ldots, L\}$ and its corresponding vector of Modified "slack" $\{S_i^M; i = 1, 2, \ldots, L\}$, we calculate step size as following:

$$\Delta = G_{fac} \times \sum_{i=1}^{L} S_i^M, i = 1, 2, \ldots, L. \tag{4.6}$$

Where G_{fac} is a global factor to adjust watermarking strength. Then we use this step size as Δ in Equation (3.32) to do STDM embedding. On the other hand in the detector, we firstly calculate modified slack according to received signal and then get $\hat{\Delta}$ in the same way as Equation (4.6). Finally, the detected bit is determined by Equation (3.33).

STDM-OptiMW-SS

Our experimental results, described next, revealed that the performance of STDM-MW-SS did not perform as well as expected with respect to JPEG compression.

To understand this, we have further investigated how the modified Watson affects the slacks. For the $1,000$ images with dimension of 768×512, slacks are computed based on 8×8 blocks. Then, an average value of the slacks for each coefficient in an 8×8 block is obtained. These average values are shown in Table 4.1 and Table 4.2 for the regular Watson's slacks and modified Watson's slacks, respectively. The ratios of modified slacks to regular slacks are shown in Table 4.3.

Table 4.3 shows that all ratios are larger than 1, which indicates that all average values of our modified perceptual model are larger than the regular Watson's slacks. Further, ignoring the slacks corresponding to the DC term which we do not actually alter, Table 4.3 illustrates that the ratios are larger for higher DCT frequency coefficients, rising from 1.72 for the lowest frequency to 5.84 for the highest frequency.

This investigation revealed that the inferior performance against JPEG of STDM-MW-SS, which is based on our modified perceptual model, was due to the fact that

1.32	13.63	7.75	6.56	5.91	5.62	5.96	6.85
11.57	8.04	5.59	4.80	4.57	4.47	4.54	5.11
6.24	5.35	5.61	5.10	4.92	4.87	5.13	5.93
4.93	4.32	4.80	5.33	5.61	5.88	6.43	7.40
4.29	3.73	4.22	5.33	6.60	7.51	8.44	9.63
4.25	3.60	4.22	5.55	7.40	9.26	10.96	12.70
4.94	3.90	4.65	6.16	8.32	10.93	13.61	16.20
6.31	4.81	5.67	7.21	9.59	12.68	16.19	19.81

Table 4.1: Average of regular slacks for each DCT coefficient

7.67	23.47	14.21	13.78	15.80	20.03	26.82	35.96
20.09	15.17	11.35	10.94	12.64	15.63	20.37	27.05
12.02	10.97	14.52	15.65	17.25	20.42	25.32	32.20
11.67	10.47	15.39	21.27	25.27	29.16	34.39	41.56
14.13	11.83	16.86	25.19	33.80	40.85	47.61	55.59
19.04	15.19	20.15	29.07	40.82	52.60	63.31	73.86
26.26	20.18	25.19	34.34	47.61	63.31	79.22	94.49
35.87	27.00	32.15	41.52	55.59	73.86	94.49	115.59

Table 4.2: Average of modified slacks for each DCT coefficient

5.83	1.72	1.83	2.10	2.67	3.57	4.50	5.25
1.74	1.89	2.03	2.28	2.77	3.50	4.49	5.30
1.93	2.05	2.59	3.07	3.51	4.20	4.94	5.43
2.37	2.42	3.21	3.99	4.50	4.96	5.35	5.62
3.30	3.17	4.00	4.72	5.12	5.44	5.64	5.77
4.48	4.22	4.78	5.24	5.52	5.68	5.78	5.82
5.31	5.18	5.42	5.58	5.72	5.79	5.82	5.83
5.68	5.61	5.67	5.76	5.80	5.83	5.83	5.84

Table 4.3: Ratios of modified slacks to regular slacks for each DCT coefficient

our modified perceptual model was (i) generally producing larger slack estimates than Watson's model and (ii) that this error was larger for high frequency DCT coefficients. Thus, much more of the watermark signal was being placed in the very high frequency DCT coefficients which are the first to be eliminated by JPEG compression.

To reduce this affect, we altered our perceptual model to more closely follow the original Watson model. Note that this alteration must still retain the linear scaling characteristic necessary to provide resistance to valumetric scaling. This is accomplished by creating a piecewise linear model in which the modified slacks calculated for all frequencies are divided by their corresponding ratios illustrated in Table 4.3. This new modified perceptual model is called OptiMW (Optimal Modified Watson), while the new slacks are denoted as S^{OptiM}.

Once again, these new slacks can be used not only to determined the projection vector for STDM, but also to adaptively select the quantization step size. We calculate the step size in the same way as we did in Equation (4.6) for STDM-MW-SS, except only replacing the modified slacks, S^M, with these optimal modified slacks, S^{OptiM}. This results in a new method, referred to as STDM-OptiMW-SS.

Chapter 5

Experimental Results

We have introduced previous QIM methods in Chapter 3 and proposed our QIM algorithms based on Watson's model in in Chapter 4. This chapter is to demonstrate the experimental results and compare the performances of the methods discussed in Chapters 3 and 4.

5.1 Comparison between mean square error and Watson's perceptual distance

To contrast Watson's perceptual distance with measures of fidelity based on DWR or PSNR, Figure 5.2(a) and 5.2(b) shows two images with a DWR=15dB and a PSNR of 28dB, and the original image is shown in Figure 5.1.

Clearly, Figure 5.2(b) has a much higher fidelity that Figure 5.2(a), yet the DWR and PSNR measure are identical. In contrast, the Watson distances for Figures 5.2(b) and 5.2(a) are 53 and 513, respectively. Figure 5.3(a) and Figure 5.3(b) shows that same images at a DWR of 35dB and a PSNR or 49. The Watson distances of Figure 5.3(b) and Figure 5.3(a) are 8 and 39 respectively. It is very difficult to discern any difference even viewed on a high resolution computer monitor. However, if the images are magnified there are subtle artifacts within the sky region.

Finally, we note that the Watson distance is not normalized by the size of the image, while the DWR and PSNR measure are. Consequently, a Watson distance of say, 100, is substantially worse for small images than for large images.

Figure 5.1: Original Image

To get better compression methods for images, people need to find optimal balance point between higher compression rate and lower perceptual artifact. Similarly, to obtain better digital watermarking methods, it is required to develop more elaborate compromise between robustness and fidelity of watermarked images. For both demands, it is very important and desirable to investigate visual perceptual models. The Watson's model introduced in this section is used later to design watermarking schemes.

5.2 Experimental Results for the Adaptive Method Based on Regular Watson Model

We used a database of 1,000 images from the Corel database, each of dimension 768×512. A binary message of length $12,288$ bits is embedded into each image. We extracted 62 DCT coefficients from each 8×8 block, ignoring the DC and highest frequency coefficients. The entire sequence of 62×6144 coefficients were then pseudo-randomized and each bit of the message was embedded in 31 random coefficients. This

(a) Image with a Watson distance of 513, DWR= 15dB and PSNR= 28dB

(b) Image with a Watson distance of 53, DWR= 15dB and PSNR= 28dB

Figure 5.2: Processed images, DWR = 15dB

(a) Image with a Watson distance of 39, DWR= 35dB and PSNR= 49dB

(b) Image with a Watson distance of 8, DWR= 35dB and a PSNR= 49dB

Figure 5.3: Processed images, DWR = 35dB

is equivalent to embedding two bits in each block of the image.

The same test database and embedding rate is used for all the experimental results shown in this chapter except some test on standard images that have a different dimension. That is illustrated in Section 5.6.

Randomization of the coefficients ensures that the 31 coefficients associated with a single bit are (i) distributed spatially throughout the image, and (ii) distributed across a variety of low, mid and high frequency coefficients. This provides some robustness to (i) clipping and other spatially localized processing and (ii) frequency filtering. The randomization also has the effect of whitening the host signal.

To see the benefit that randomization of the host signal can bring, we examined the watermarking effectiveness of the embedder with and without randomization of the host coefficients. The effectiveness is simply the BER when detection is done to the generated watermarked signal immediately after embedding, i.e. in the absence of any subsequent distortion. Given the experimental conditions described early in this section, it is observed that BER of effectiveness with Non-Randomization method is 0.042. In contrast, BER with method of randomization cross whole image is just 0.003. This demonstrates over an order of magnitude improvement with randomization.

Figure 5.4 shows the bit-error-rate (BER) as a function of additive white Gaussian noise. Results are provided for both our adaptive method and the original DM algorithm of [BG99b]. In both cases, we adjusted the watermark strength such that the DWR=35dB.

To fix the average DWR at 35dB, we fixed the quantization step size at 3.0 for DM. For adaptive DM, the global constant, G, was adjusted to a value of 0.3 to meet the requisite DWR. Each point on a curve is the BER averaged over 1000 images from the Corel database.

The original DM algorithm has superior performance for low noise. We believe the poorer performance of the adaptive DM algorithm in the low noise region is due to a combination of (i) discrepancies between the corresponding estimated quantization step sizes at the embedder and the decoder, and (ii) the smaller step sizes used.

5.2. Experimental Results for the Adaptive Method Based on Regular Watson Model93

For the adaptive DM method based on Watson's model, the histogram of quantization step sizes used for each image is shown in Figure 5.5. The overall average for 1,000 images is 2.415, which is smaller than that used for standard DM. Note that the step sizes used for both methods quantize DCT coefficients, while the noise is added in the pixel domain of image.

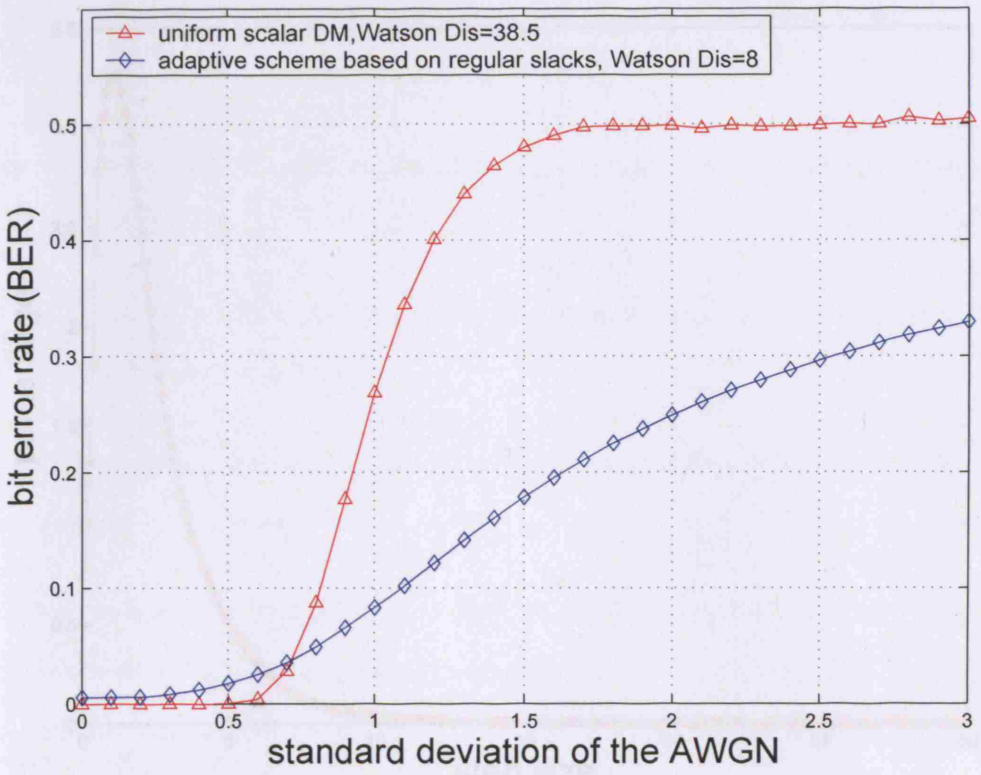


Figure 5.4: Bit error rate as a function of additive white Gaussian noise (DWR=35dB)

When the standard deviation in the noise exceeds 0.7, which is roughly $\Delta/4$ ($\Delta = 3.0$), the BER for DM rapidly degrades, and the adaptive method is clearly superior. Note also, that the superior performance of our algorithm is achieved with a very low Watson distance of 8 (i.e. very high fidelity) compared with the original method which has a Watson distance of 38.5. Thus, improved robustness and improved fidelity have been simultaneously achieved.

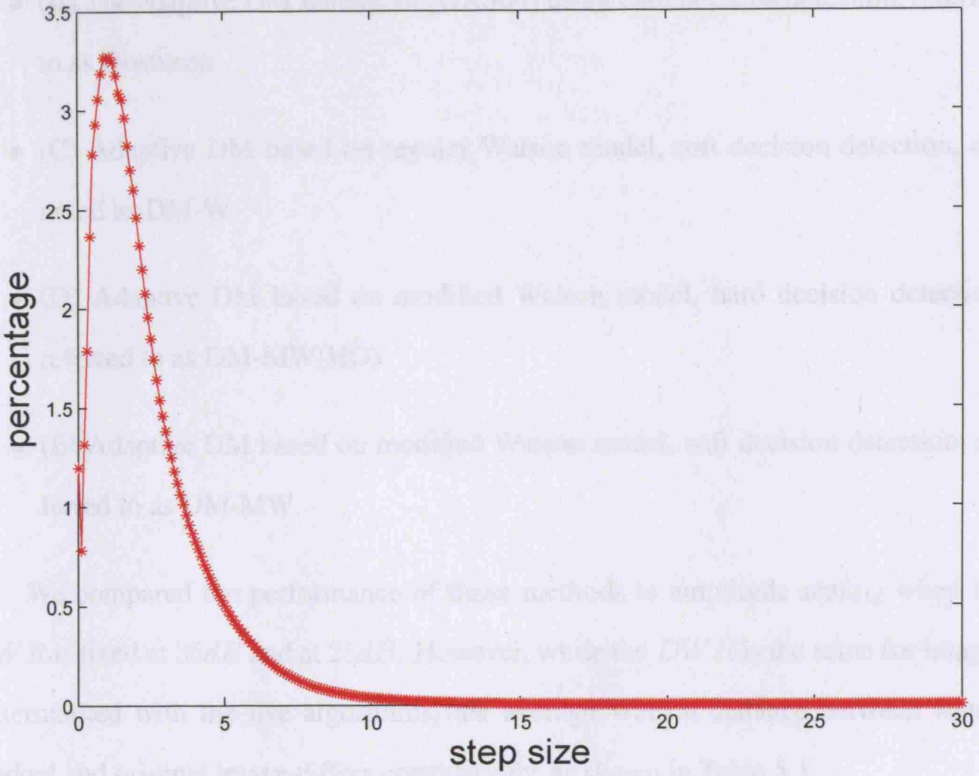


Figure 5.5: Histogram of step size for adaptive DM using Watson's model. The average step size for 1,000 image is 2.415

5.3 Experimental Results for DM Using Modified Watson

For comparison purposes, we evaluated the performance of the following algorithms:

- (A) The original non-adaptive DM scheme of [BG01b] using soft decision detection, denoted as DM
- (B) The adaptive DM scheme of [OKS04] using hard decision detection, referred to as Oostveen
- (C) Adaptive DM based on regular Watson model, soft decision detection, denoted as DM-W
- (D) Adaptive DM based on modified Watson model, hard decision detection, referred to as DM-MW(HD)
- (E) Adaptive DM based on modified Watson model, soft decision detection, referred to as DM-MW

We compared the performance of these methods to amplitude scaling when the DWR is fixed at $35dB$ and at $25dB$. However, while the DWR is the same for images watermarked with the five algorithms, the average Watson distance between watermarked and original image differs considerably, as shown in Table 5.1.

Scheme	Watson Distance
(A) DM	38.5
(B) Oostveen	55.2
(C) DM-W	8
(D) DM-MW(HD)	9.4
(E) DM-MW	9.4

Table 5.1: Average Watson distance for different methods.

Table 5.1 shows that the three adaptive schemes proposed here have very much lower perceptual distortion as measured by Watson distance. Importantly, the modification to the Watson distance used in methods D and E to provide robustness against

valumetric scaling, produces only a small degradation in image quality and remain much better than methods *A* or *B*.

The robustness to amplitude scaling for all schemes is shown in Figures 5.6 and 5.7 for DWR's of 35dB and 25dB respectively. The performance is qualitatively similar for both DWR's, though, of course, the BER is considerably smaller for DWR = 25dB, since the watermark is stronger in this case. The discussion below is therefore restricted to the case of DWR = 35dB.

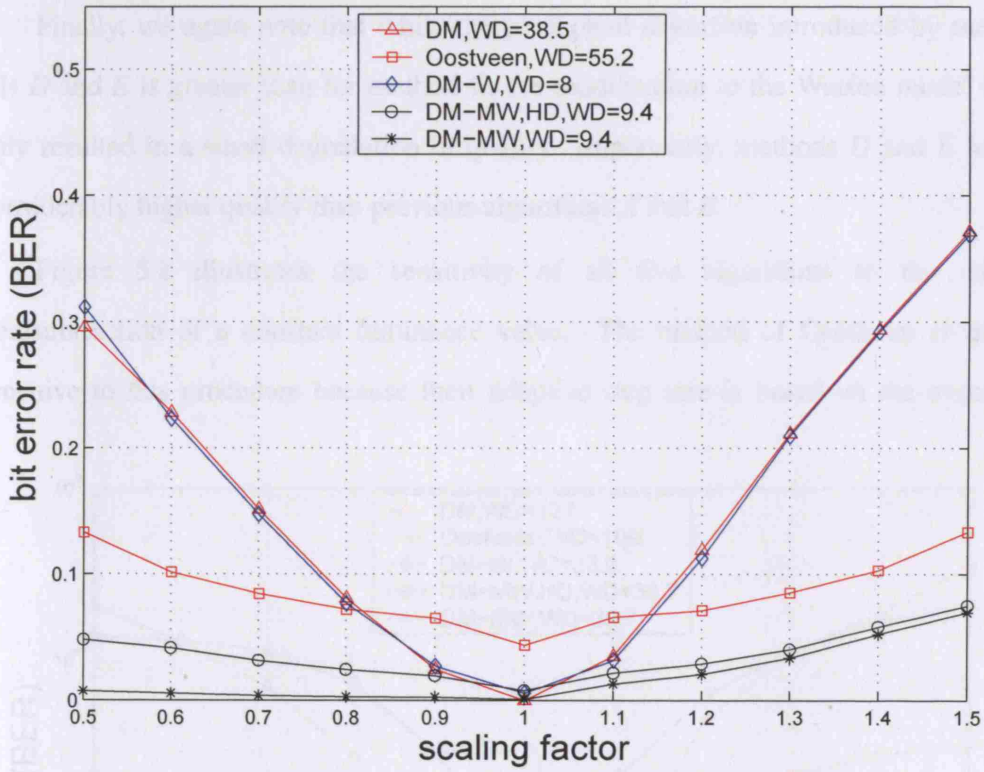


Figure 5.6: BER versus amplitude scaling (DWR = 35dB)

We observe that for very small changes in scale, $0.9 \leq \beta \leq 1.1$, the original algorithm *A* performs as well or better than the others. Our method *C* has poorer performance in this range, but for larger scale changes, it has similar or superior performance. It is also important to note that this is achieved with a perceptual distortion, as measured by Watson distance, of less than 20% compared with method *A*, (see Table 5.1).

Both algorithms *A* and *C* are not designed to be invariant to valumetric scaling. Bit error rates of greater than 10% occur for $\beta < 0.8$ and $\beta > 1.1$. In contrast, Oostveen *et al*'s method *B* and our methods *D* and *E* show much better robustness to scale changes.

Clearly method *E* outperforms all others with a BER that never exceeds 7% over the range of β tested. To ensure that this performance was not due to soft decoding alone, we implemented method *D*, DM-MW with hard decision. While performing worse than method *E*, method *D* is still superior to Oostveen *et al*'s method (which also uses hard decoding). One possible explanation for the poorer performance of Oostveen *et al*'s method is that it is pixel-based rather than DCT-based. Consequently, clipping of the pixels at 0 and 255 may have a more direct and deleterious effect.

Finally, we again note that while the perceptual distortion introduced by methods *D* and *E* is greater than for method C, the modification to the Watson model has only resulted in a small degradation in quality. Importantly, methods *D* and *E* have considerably higher quality than previous algorithms *A* and *B*.

Figure 5.8 illustrates the sensitivity of all five algorithms to the addition/subtraction of a constant luminance value. The method of Oostveen *et al* is sensitive to this procedure because their adaptive step size is based on the average

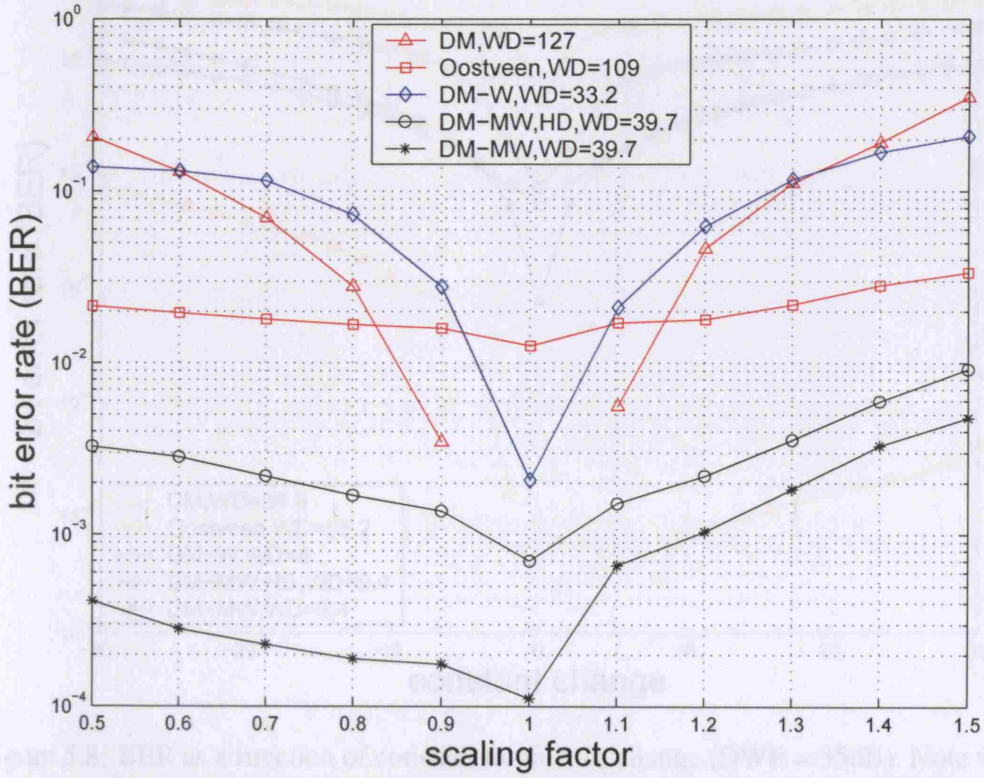


Figure 5.7: BER versus amplitude scaling, (DWR = 25dB). Note that for basic DM, the BER when there is no (unity) scaling is 0, and this point is therefore not plotted.

intensity of pixels in a neighborhood. This is clearly demonstrated in Figure 5.8, where it has the highest bit error rate. Conversely, the regular DM algorithm performs best, as the quantization occurs in the DCT domain and does not include the DC coefficient. Thus, we expect that DM would not be affected by this operation. Nevertheless, performance degradation is observed, particularly as we darken (subtract) the image. We suspect that this is due to clipping artifacts. Our proposed algorithms are substantially worse than DM, but about an order of magnitude less sensitive compared with Oostveen *et al.* The sensitivity of our methods is due to that fact that the slack calculations are based on the average intensity of the image, see Equation (4.3), which is altered by the addition of a constant luminance.

Figure 5.9 shows the BER as a function of amplitude scaling after the addition of a constant luminance change of 10. The performance of all algorithms is qualitatively similar. However, the BER is, on average, considerably worse than for amplitude

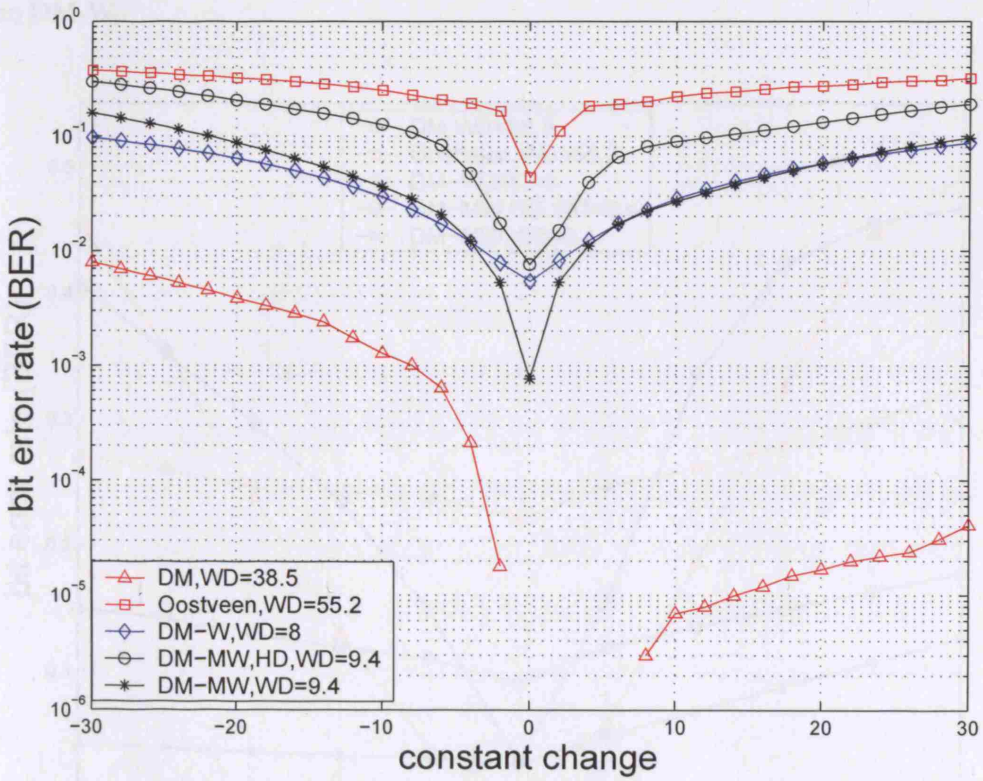


Figure 5.8: BER as a function of constant luminance change (DWR = 35dB). Note that for basic DM, the BER when there is no change in luminance is 0, and this point is therefore not plotted.

scaling alone, as shown in Figure 5.6 .

Figure 5.10 shows the sensitivity of all five algorithms to additive, white, Gaussian noise. The two curves from Figure 5.4 are included for completeness. All of the adaptive methods perform worse than regular DM, for low levels of noise. However, the adaptive methods degrade more gradually as the standard deviation of the noise increases above about 0.6. The best performing algorithms in this high noise regime are DM-MW using soft decoding and DM-W. Interestingly, the DM-W demonstrates the best performance for noise with standard deviations of greater than 1.7.

To understand why the performance of DM-W outperforms DM-MW for noise of standard deviation greater than about 1.7, we examined the cumulative distribution in quantization step sizes for the two algorithms, as shown in Figure 5.11. The two curves diverge at a step size of about 1.8. Thereafter, the curve representing DM-MW increases more quickly, indicating that the DM-MW has smaller quantization step sizes than DM-W.

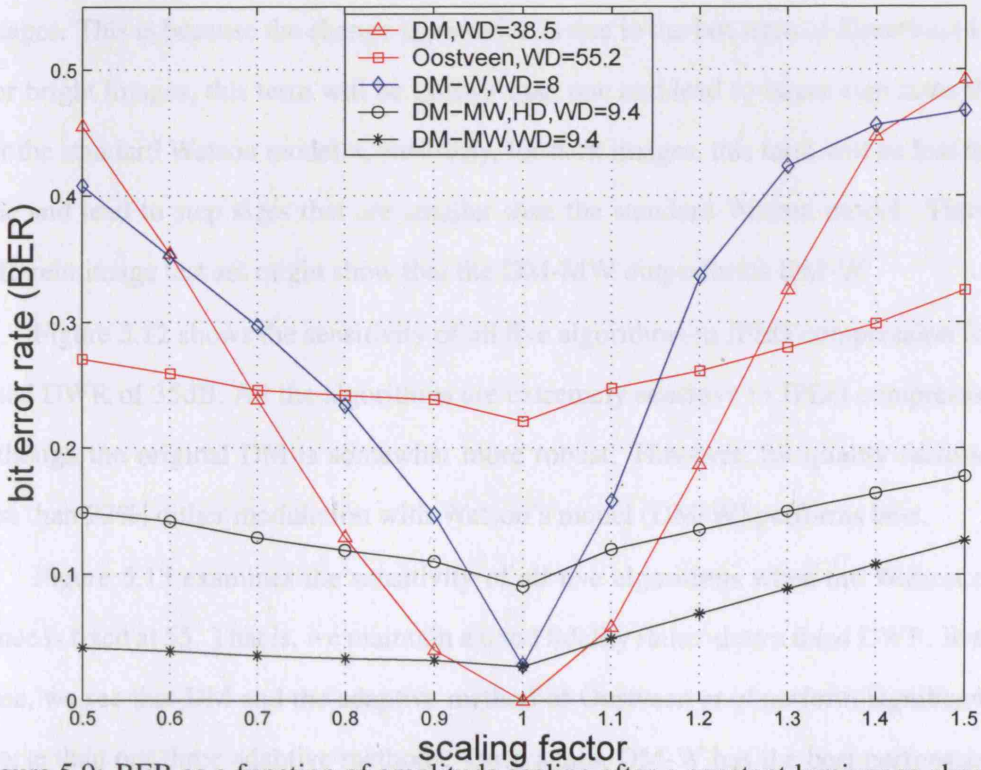


Figure 5.9: BER as a function of amplitude scaling after a constant luminance change of 10 (DWR = 35 dB)

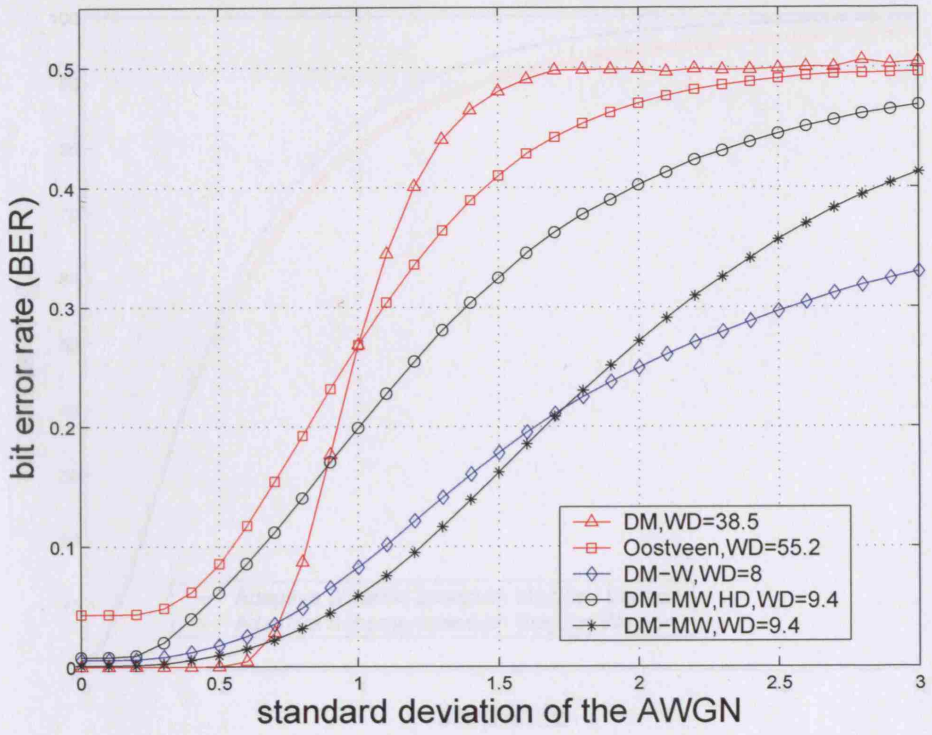


Figure 5.10: BER as a function of additive white Gaussian noise (DWR = 35dB).

It should be noted that this performance difference is specific to a particular set of images. This is because the change in step size is due to the last term of Equation (4.2). For bright images, this term will be greater than one and lead to larger step sizes than for the standard Watson model. Conversely, for dark images, this term will be less than one and lead to step sizes that are smaller than the standard Watson model. Thus, a different image test set might show that the DM-MW outperforms DM-W.

Figure 5.12 shows the sensitivity of all five algorithms to JPEG compression for a fixed DWR of 35dB. All the algorithms are extremely sensitive to JPEG compression, although the original DM is somewhat more robust. However, for quality factors of less than 92%, dither modulation with Watson's model (DM-W) performs best.

Figure 5.13 examines the sensitivity of all five algorithms when the Watson distance is fixed at 55. That is, we maintain a fixed fidelity rather than a fixed DWR. In this case, we see that DM and the adaptive method of Oostveen *et al* perform significantly worse than our three adaptive methods. Once again, DM-W has the best performance for JPEG quality factors less than 75.

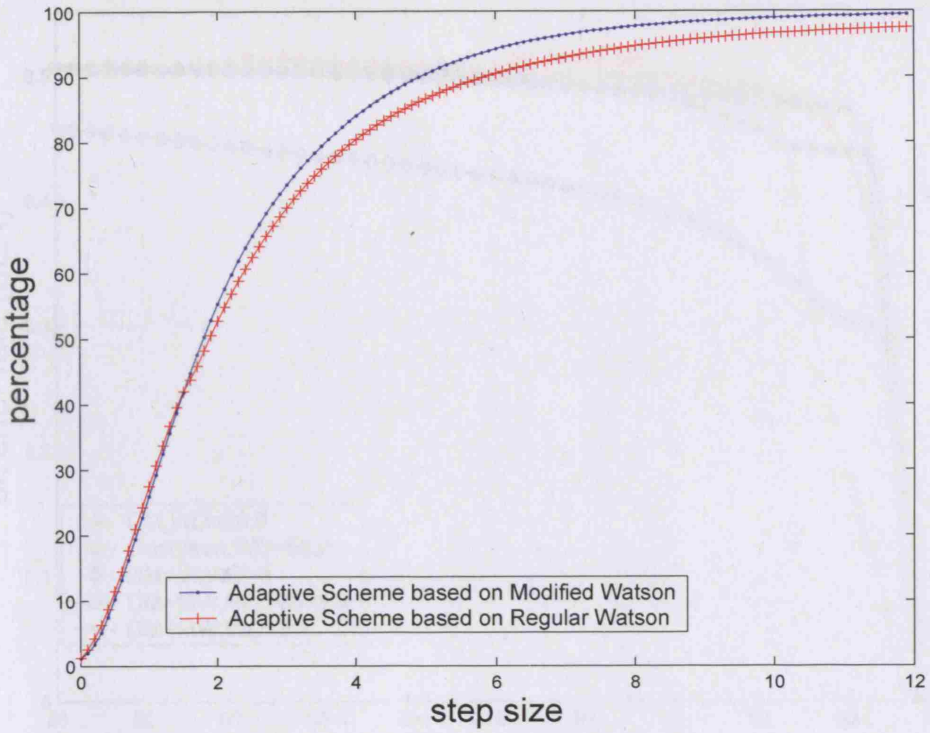


Figure 5.11: Cumulative histogram of quantization step sizes for DM-WM and DM-M, (measured over all 1000 images).

5.3.1 Performance on sample images from different categories

Figure 5.6 shows the average experimental results derived from 1,000 images from the Corel database. To make up these 1,000 test images, 20 categories of Corel database are used and 50 images are selected form each category. We now further investigate the performance of our method based modified Watson by examining various image samples form different categories. For example, Figures 5.14(a), 5.15(a) are from the category of "Wild animals", Figures 5.16(a), 5.17(a) are from the category of "Greece". They are all watermarked images using DM-MW. Figures 5.14(b), 5.15(b), 5.16(b) and 5.17(b) are their versions after amplitude scaling. It is clearly seen that all images become obviously brighter after scaling up by a factor of 1.5.

Mathematically, our modified Watson is designed to scale linearly with amplitude scaling of pixel values. Consequently, the step size used in DM-MW should also be scaled linearly with the scaling factor. In the end, the BER should be theoretically

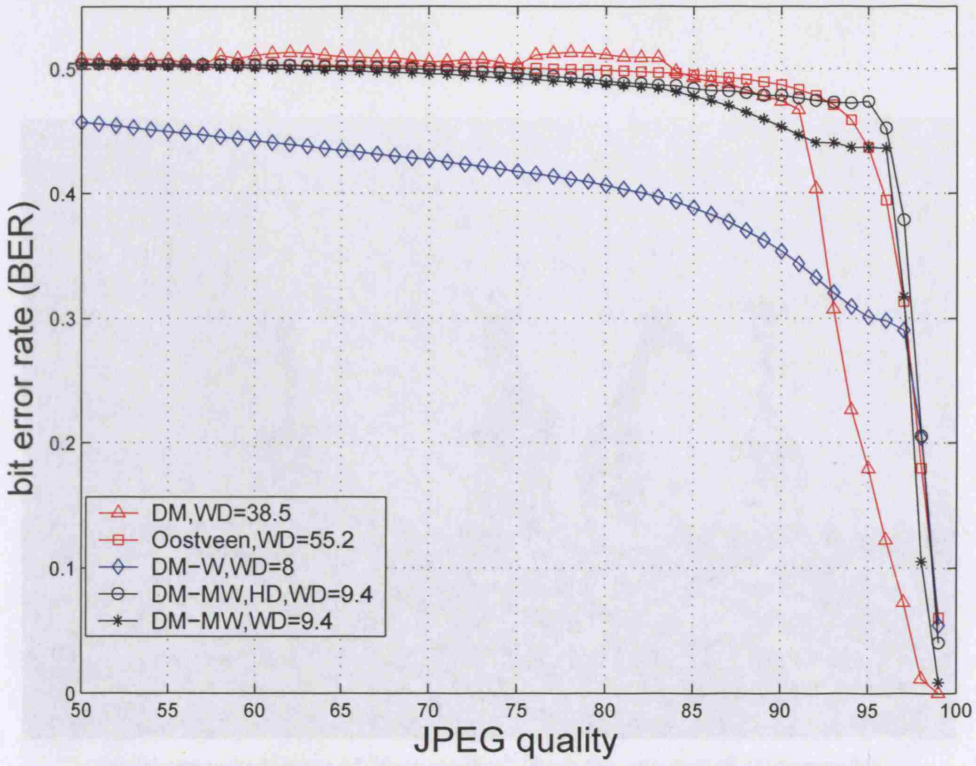


Figure 5.12: BER as a function of JPEG quality for a fixed DWR of 35dB.

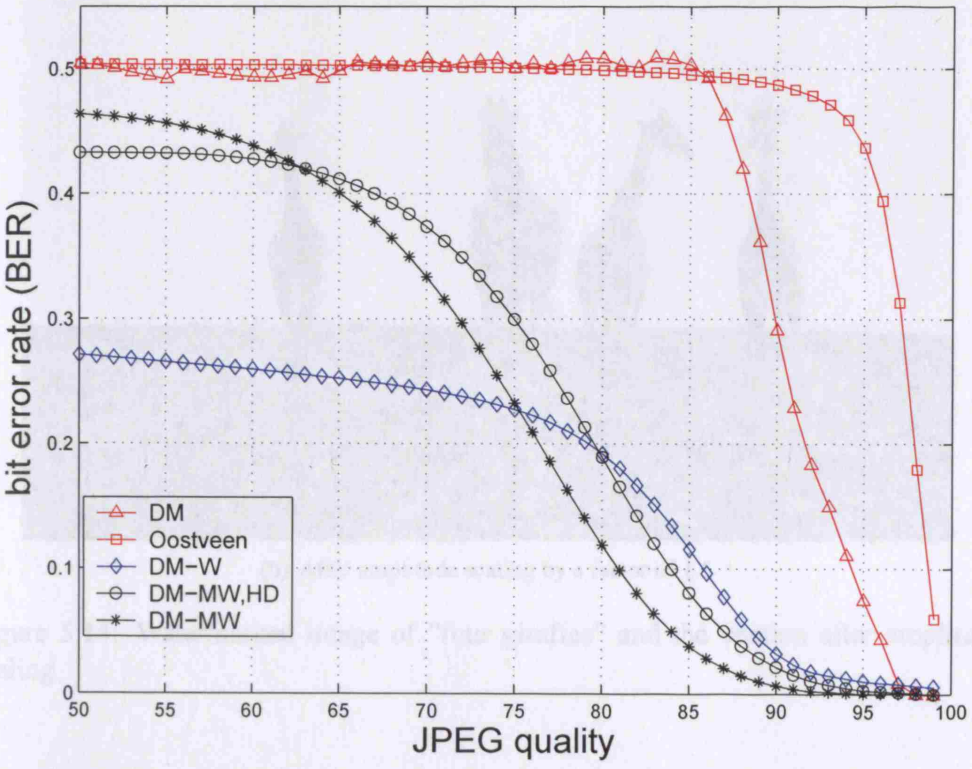


Figure 5.13: BER as a function of JPEG quality for a fixed Watson distance of 55.

(a) Watermarked image of "four giraffes" (from the category of wild animals)

(b) After amplitude scaling by a factor of 1.5

Figure 5.14: Watermarked image of "four giraffes" and the version after amplitude scaling

(a) Watermarked image of "drinking giraffe" (from the category of wild animals)

(b) After amplitude scaling by a factor of 1.5

Figure 5.15: Watermarked image "drinking giraffe" and the version after amplitude scaling

(a) Watermarked image of "White building" (from the category of Greece)

(b) After amplitude scaling by a factor of 1.5

Figure 5.16: Watermarked image "White building" and the version after amplitude scaling

(a) Watermarked image of "Buildings" (from the category of Greece)

(b) After amplitude scaling by a factor of 1.5

Figure 5.17: Watermarked image "White building" and the version after amplitude scaling

invariant to either scaling factor or image pixel values. However, the BER differences between various images are considerable. Table 5.2 illustrates the BER detected from the four scaled version of watermarked images.

Figure Index	5.14(b)	5.15(b)	5.16(b)	5.17(b)
BER	0.272	0.018	0.267	0.017

Table 5.2: BER of various watermarked images after scaling up by a factor of 1.5

Given the same experimental conditions, Figures 5.14(b) has a high BER of 0.272 while Figure 5.16(b) has a BER as low as 0.018, although they are from a same image category named "Wild animals". Similarly, the BERs of another two images form the category of "Greece" are compared. Figures 5.16(b) and 5.17(b) have BER of 0.267 and 0.017, respectively. From further investigation, we understand the considerable differences on BER are not because of the faults of our modified Watson method, but instead, because of the limitations of the test image format. Please note that all test images in this thesis are in the 256 grayscale format, which means the pixel value can only be integers ranging from 0 to 255. Imagine that a pixel value is larger than 170 and if it is scaled up by a factor of 1.5, the generated pixel needs to be floored to 255. In these cases, the image generated by amplitude scaling is not strictly the "scaled" version of the watermarked image. Our modified Watson system may break down if these are too many pixels generated by the floor function described above. Obviously, both Figure 5.14(b) and 5.16(b) have large area of pixels floored to 255. This explains why they have a much higher BER than Figures 5.15(b) and 5.17(b). In fact Figures 5.14(b) and 5.16(b) are among the worst performance images in the database.

We haven't found a clear link the categories (contents) of the images and the performances. The images from the same category may have quite different performances, as the sample images illustrated in this section. We examined the average BER in several categories as plotted in Figure 5.18. It shows that average BER of 50 images is rather stable and there are no considerable differences between categories.

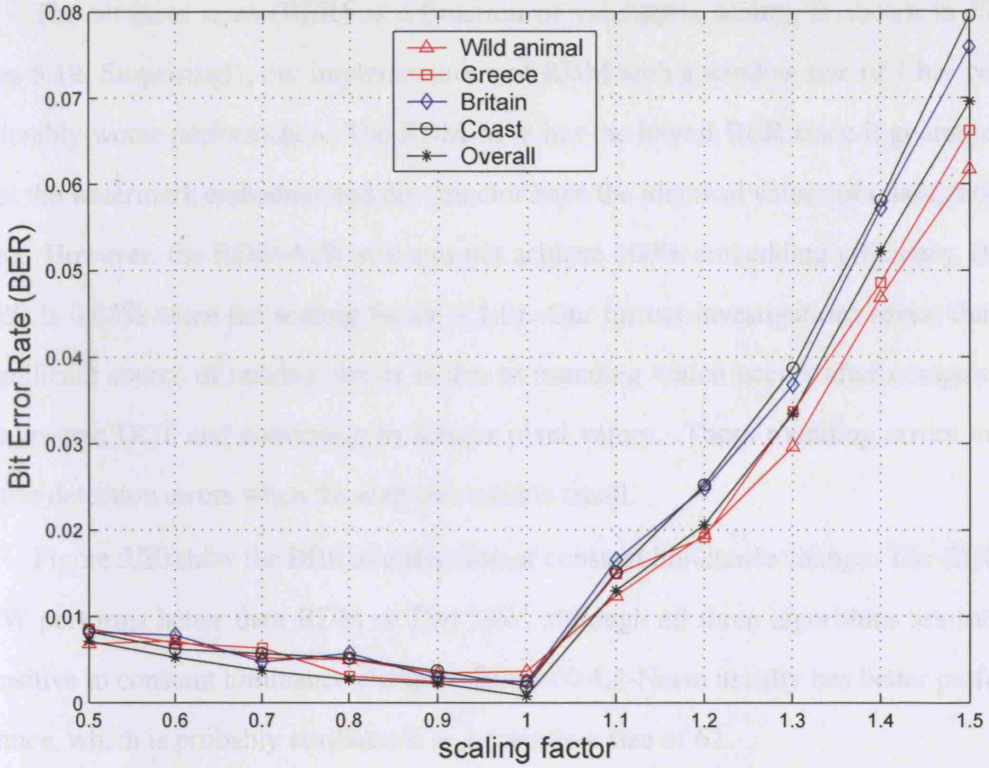


Figure 5.18: BER versus amplitude scaling, various categories

5.4 Experimental Results for RDM using Modified Watson

Once again, we watermarked 1,000 images from the Corel database. And in all experiments, the average document-to-watermark ratio (DWR) was fixed at 35dB.

The average Watson distance of the watermarked images for each of the three RDM algorithms and DM-MW is tabulated in Table 5.3. As expected, RDM-MW compared with DM-MW (9.4). However, the difference is small. RDM has a much larger Watson distance of 54 and RDM-62-L2-Norm has a perceptual distance of 29. Clearly, RDM-MW has provided a much reduced perceptual distortion.

Scheme	Watson Distance
DM-MW	9.4
RDM-MW	10.2
RDM	54
RDM-62-L2-Norm	29

Table 5.3: Average Watson distance for various methods.

The bit error rates (BER) as a function of valumetric scaling is shown in Figures 5.19. Surprisingly, our implementation of RDM with a window size of 1 has considerably worse performance. The RDM-MW has the lowest BER since it guarantees that the watermark embedder and the detector have the identical values of quantization step. However, the RDM-MW still can not achieve 100% embedding efficiency, (the BER is 0.03% when the scaling factor is 1.0). Our further investigations reveal that a significant source of residual errors is due to rounding which occurs after computing the inverse DCT and converting to integer pixel values. These rounding errors may cause detection errors when the step size used is small.

Figure 5.20 show the BER as a function of constant luminance change. The RDM-MW performs better than RDM or DM-MW, although all three algorithms are more sensitive to constant luminance changes. RDM-62-L2-Norm usually has better performance, which is probably attributable to its window size of 62.

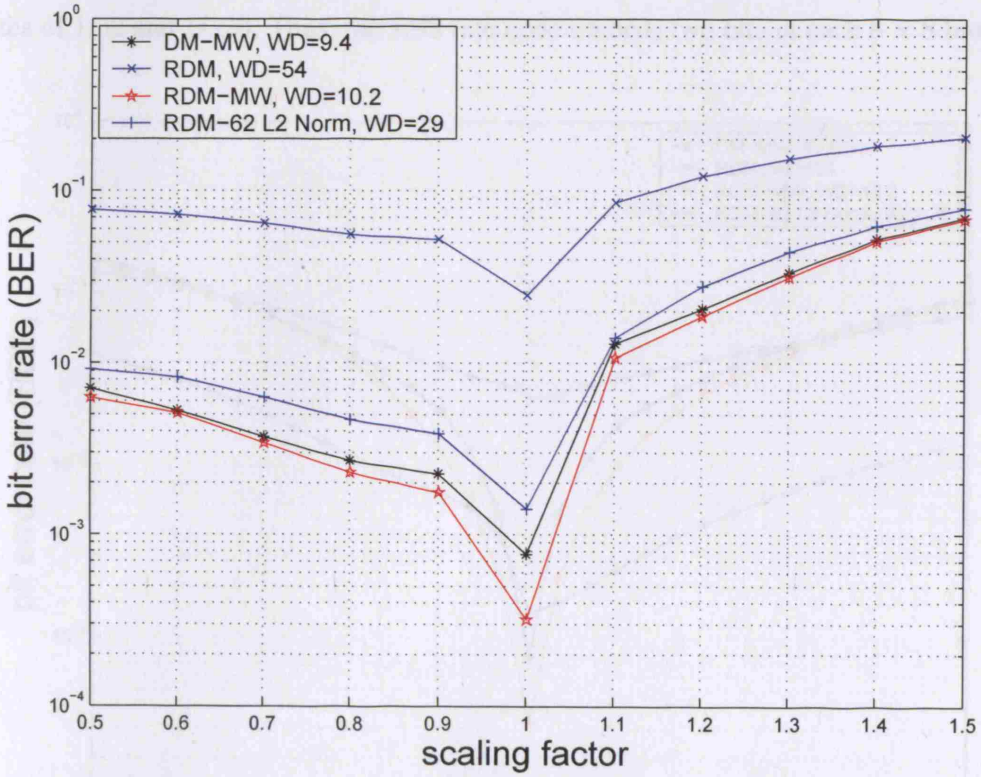


Figure 5.19: BER as a function of valumetric scaling (DWR = 35dB).

Figure 5.21 illustrates the response to additive white Gaussian noise. The RDM-

MW, RDM-62-L2-Norm and DM-MW all perform very similarly. Once again, our implementation of RDM with a window size of 1 performs significantly worse.

The sensitivity to JPEG compression is investigated in Figures 5.22 and 5.23. For a fixed DWR of 35dB we see that both RDM-MW and DM-MW perform worse than RDM algorithms with no perceptual model. However, if we fix the perceptual distortion, rather than the DWR, then the perceptual-based algorithms have superior performance. The RDM-MW has slightly worse performance than DM-MW, which is probably due to the fact that the perceptual model for RDM-MW is worse.

5.5 Experimental Results for STDM Based Methods with a Perceptual Model

Experiments are still performed on the 1000 images from the Corel image database and same set of DCT coefficients are chosen. In this section, we considered two embedding rates of 1/32 and 1/320. Thus, the 1/32 rate code embeds two bits in each 8×8 block.

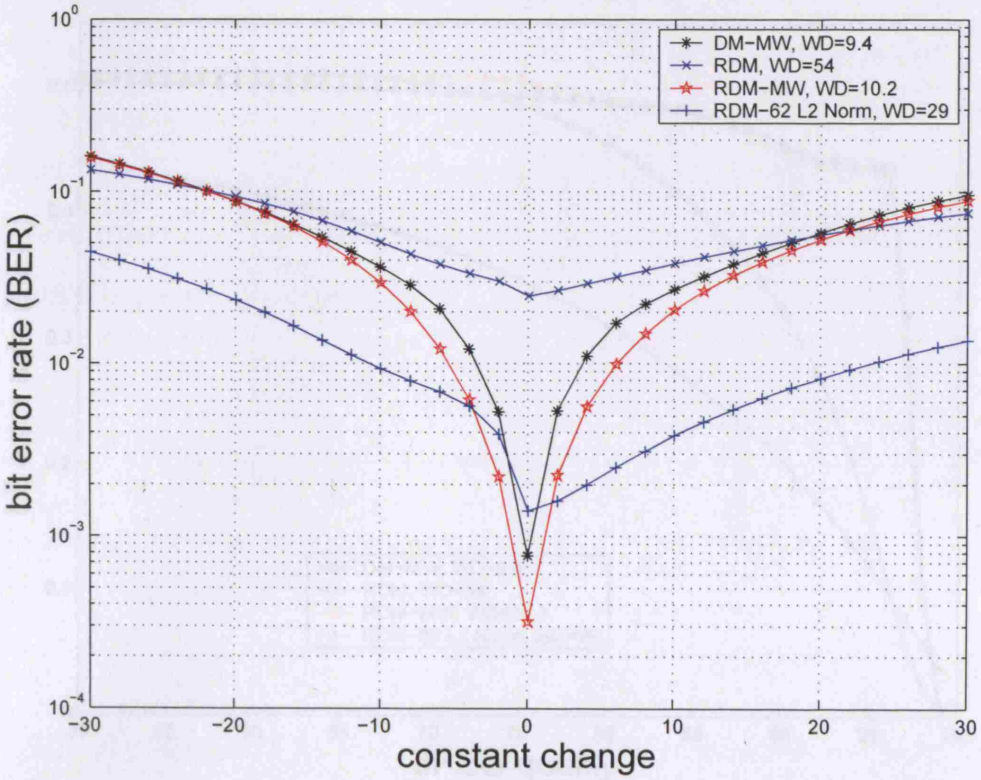


Figure 5.20: BER as a function of constant luminance change (DWR = 35dB).

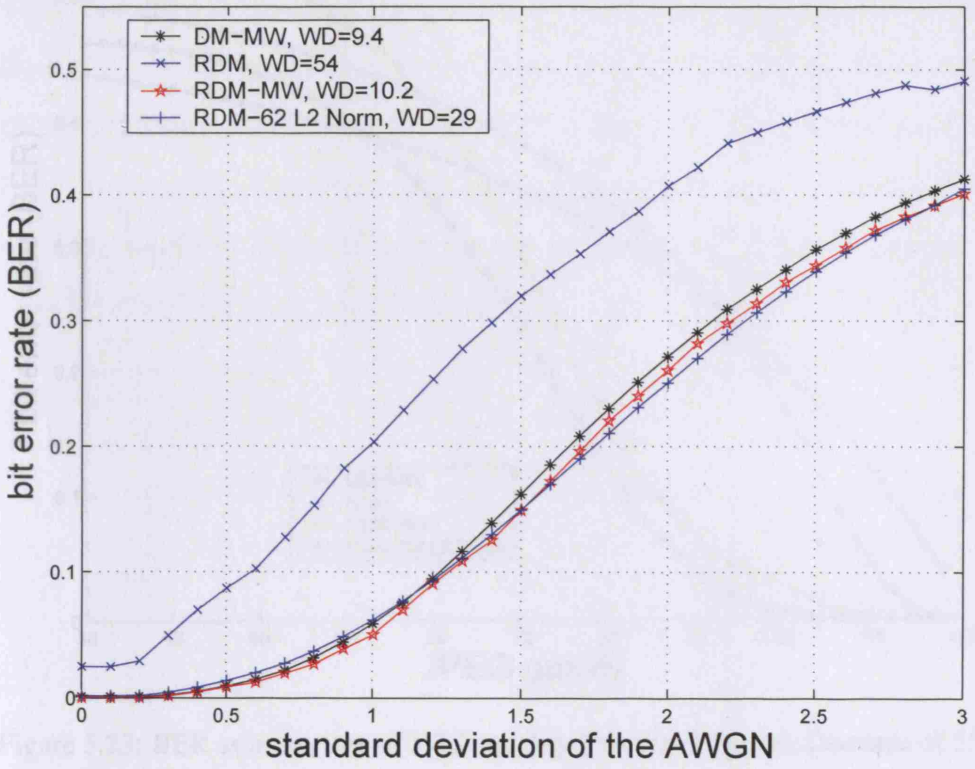


Figure 5.21: BER as a function of additive white Gaussian noise (DWR = 35dB).

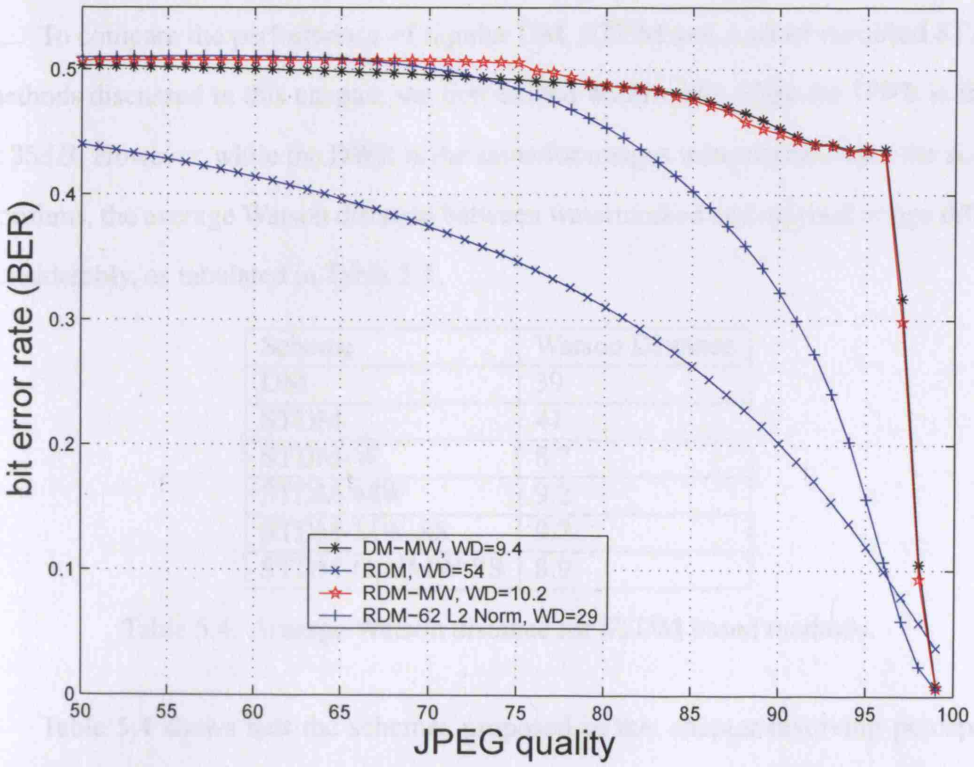


Figure 5.22: BER as a function of JPEG quality for a fixed DWR of 35dB.

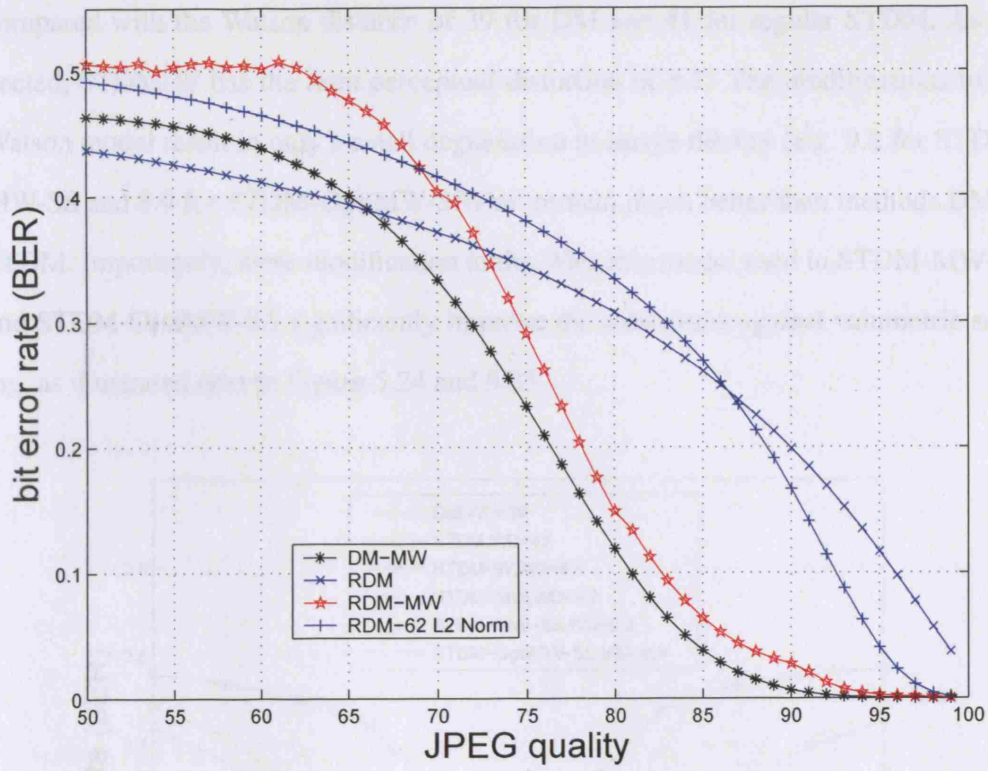


Figure 5.23: BER as a function of JPEG quality for a fixed Watson Distance of 55.

One bit is embedded into 5 blocks for the rate 1/320 code.

To compare the performance of regular DM, STDM and a set of modified STDM methods discussed in this chapter, we first embed watermarks when the DWR is fixed at $35dB$. However, while the DWR is the same for images watermarked with the six algorithms, the average Watson distance between watermarked and original image differs considerably, as tabulated in Table 5.4.

Scheme	Watson Distance
DM	39
STDM	41
STDM-W	8.7
STDM-MW	9.2
STDM-MW-SS	9.2
STDM-OptiMW-SS	8.9

Table 5.4: Average Watson distance for STDM based methods.

Table 5.4 shows that the schemes proposed in this chapter involving perceptual models have very much lower perceptual distortion as measured by Watson distance,

compared with the Watson distance of 39 for DM and 41 for regular STDM. As expected, STDM-W has the least perceptual distortion of 8.7. The modifications to the Watson model result in only a small degradation in image fidelity (e.g. 9.2 for STDM-MW-SS and 8.9 for STDM-OptiMW-SS) but remain much better than methods DM or STDM. Importantly, these modification to the Watson's model used in STDM-MW-SS and STDM-OptiMW-SS significantly improve the robustness against valumetric scaling, as illustrated next in Figure 5.24 and 5.25.

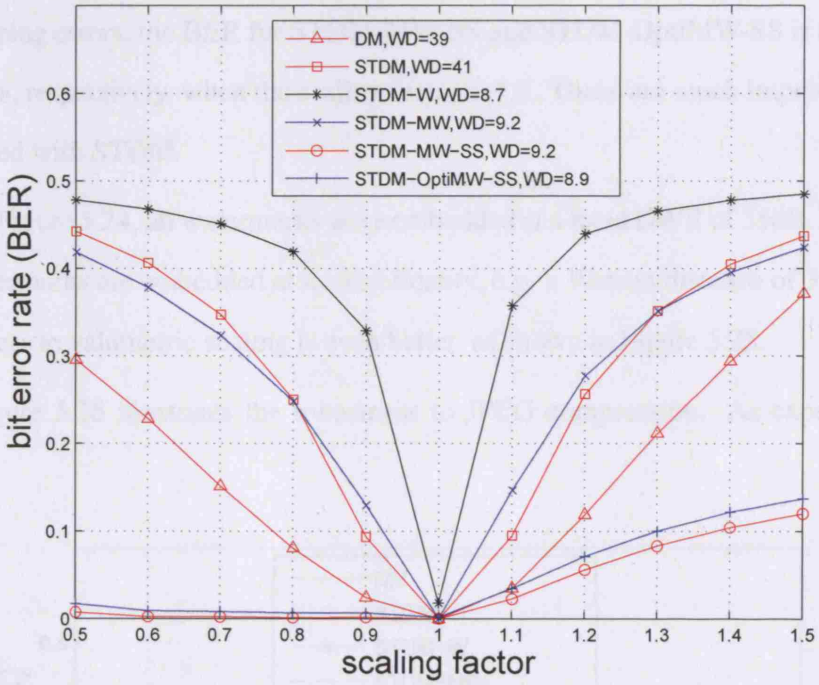


Figure 5.24: Bit error rate (BER) vs. valumetric scaling using an embedding rate of 1/32 and at a fixed DWR of 35 dB

Figure 5.24 shows the bit error rate (BER) as a function of valumetric scaling for DM, STDM and the set of modified STDM algorithms when DWR is fixed to 35 dB. As expected, DM, STDM, STDM-W and STDM-MW are sensitive to valumetric distortion. BER's for these methods are all above 30% when the scaling factor is 0.5 or 1.5. However, notice that the perceptual distortions, as measured by Watson distance, are 39, 41, 8.7 and 9.2 respectively. That is, both STDM-W and STDM-MW provide a significant improvement in fidelity. Further, the perceptual degradation due to our modified perceptual model is slight. Both STDM-MW-SS and STDM-OptiMW-

SS demonstrate very good robustness to valumetric scaling together with improved fidelity. These two methods provide very low BER rate (less than 2%), when the watermarked images are scaled down, i.e. the scaling factor is less than 1.0. Note that for STDM-MW-SS and STDM-OptiMW-SS, the robustness to amplitude scaling up (i.e. when the scaling factor is larger than 1.0) is much worse than the robustness to scaling down. This is because, in addition to scaling and rounding the pixel values to integers, amplitude scaling up introduces some clipping errors, i.e., all pixel values above 255 after scaling are truncated to 255. This has a more severe impact than rounding. Given the clipping errors, the BER for STDM-MW-SS and STDM-OptiMW-SS is about 11% and 12%, respectively, when the scaling factor is 1.5. These are much improved results compared with STDM.

In Figure 5.24, all watermarks were embedded at a fixed DWR of 35dB. If, instead, the watermarks are embedded at a fixed fidelity, e.g. a Watson distance of 39, then the robustness to valumetric scaling is even better, as shown in Figure 5.25.

Figure 5.26 illustrates the robustness to JPEG compression. As expected, DM

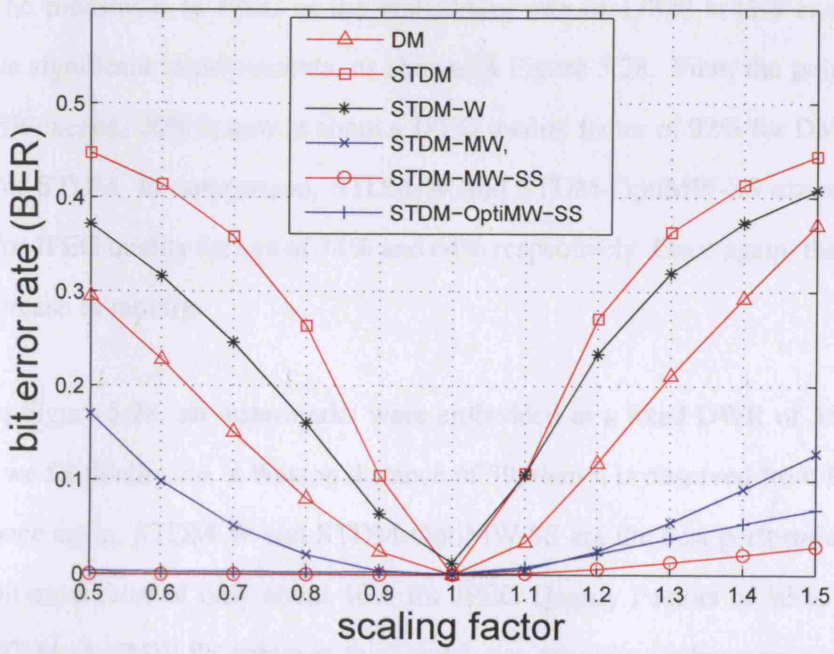


Figure 5.25: Bit error rate(BER) vs. valumetric scaling using an embedding rate of 1/32 and at a fixed Watson distance of 39

is the most sensitive, exceeded a 20% BER for a JPEG quality factor (QF) of about 94%. Standard STDM is considerably better, not exceeding a 20% BER until QF=90%. Both STDM-MW and STDM-MW-SS have almost identical performance that is better than DM but worse than STDM. As discussed earlier, this is due to that fact that the modified perceptual model over estimates the slacks for high frequency coefficients which are most sensitive to JPEG compression. By using the piecewise linear model, which reduces the slack values according to the ratios shown in Table 4.3, we see a very significant improvement in performance for STDM-OptiMW-SS. And this is achieved with a lower perceptual distortion of 8.9 as compared to 41 for STDM.

In Figure 5.26, all watermarks were embedded at a fixed DWR of 35dB. If, instead, the fidelity is fixed to a Watson distance of 39, then Figure 5.27 illustrates that DM and STDM perform worse than others and that STDM-W and STDM-OptiMW-SS are the best performing methods with bit error rates that never exceed 25% even for JPEG Quality Factors of 50%. Most importantly, STDM-OptiMW-SS achieves this whilst also providing robustness to valumetric scaling.

The robustness to JPEG at the embedding rate of $1/320$ is also examined, we observe significant improvements, as shown in Figure 5.28. First, the point at which the BER exceeds 20% is now at about a JPEG quality factor of 92% for DM and about 78% for STDM. In comparison, STDM-W and STDM-OptiMW-SS exceed the 20% BER for JPEG quality factors of 74% and 64% respectively. Once again, the BER does not increase as rapidly.

In Figure 5.28, all watermarks were embedded at a fixed DWR of 35dB. If, instead, we fix fidelity, i.e. a Watson distance of 39, then it is observed from Figure 5.29 that, once again, STDM-W and STDM-OptiMW-SS are the best performing methods with bit error rates of only about 10% for JPEG Quality Factors of 50%. Also note that STDM-OptiMW-SS achieves this whilst also providing robustness to valumetric scaling.

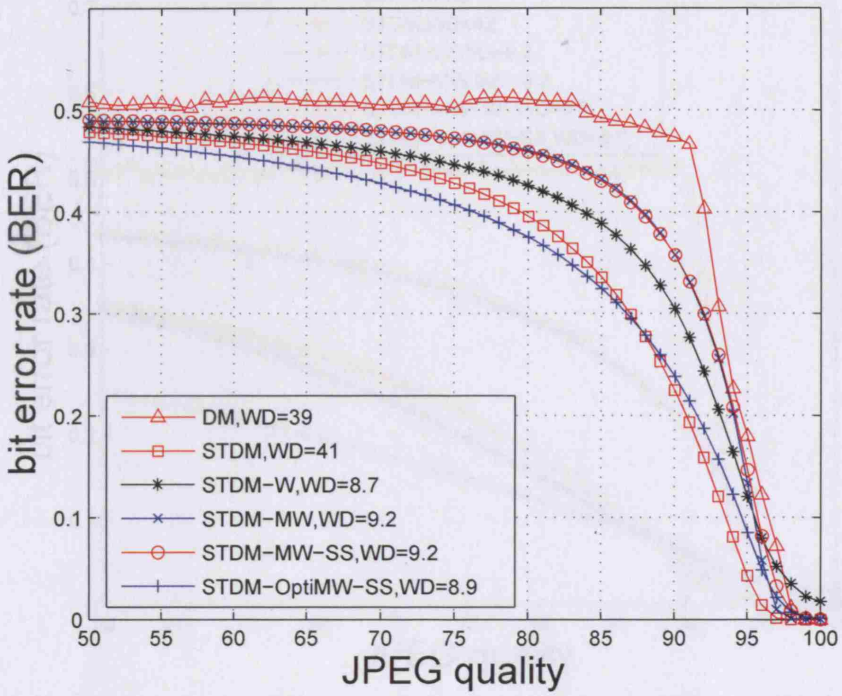


Figure 5.26: Bit error rate(BER) vs. JPEG Compression using an embedding rate of 1/32 and a DWR of 35 dB

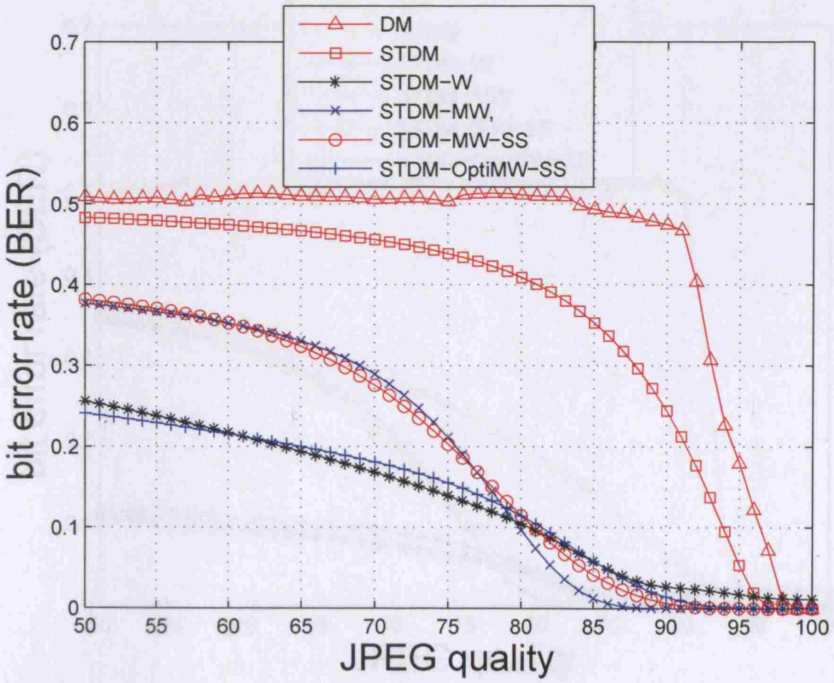


Figure 5.27: Bit error rate(BER) vs. JPEG Compression using an embedding rate of 1/32 and at a fixed Watson distance of 39

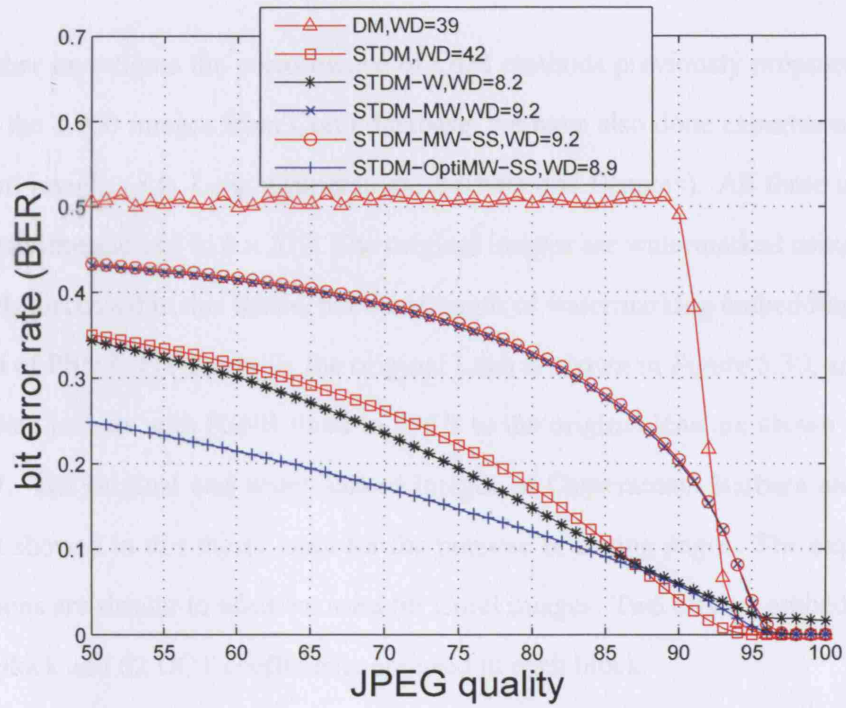


Figure 5.28: Bit error rate(BER) vs. JPEG Compression using an embedding rate of 1/320 and a DWR of 35 dB

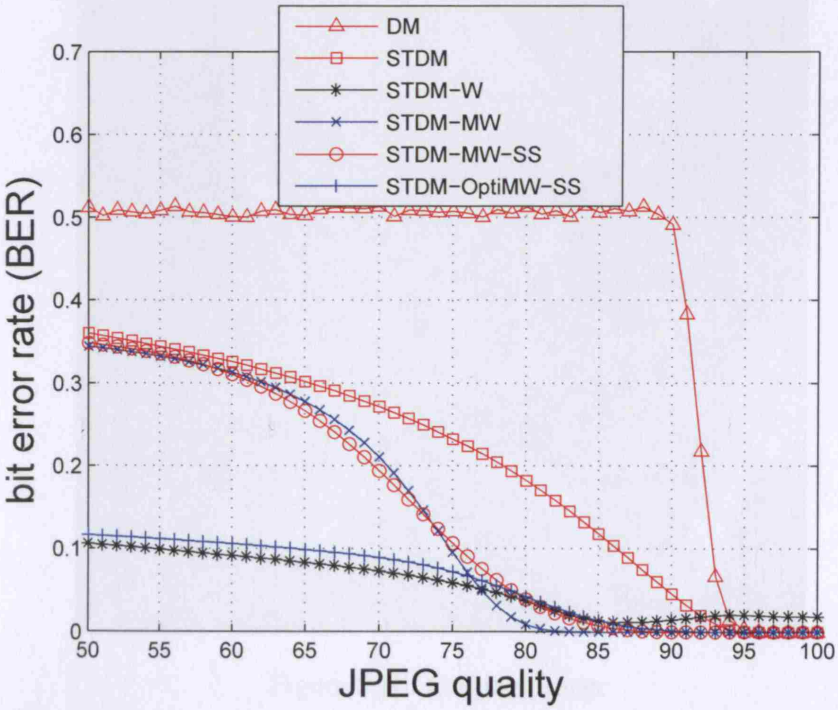


Figure 5.29: Bit error rate(BER) vs. JPEG Compression using an embedding rate of 1/320 and at a fixed Watson distance of 39

5.6 Experimental results on standard images

To further investigate the performance of QIM methods previously proposed in addition to the 1,000 images from Corel database, we have also done experiments on four standard images (e.g. Lena, Cameraman, Barbara and Peppers). All these images are with the dimension of 512×512. The original images are watermarked using different methods discussed in this thesis, but the strength of watermarking embedding are fixed in term of PSNR. For example, the original Lena is shown in Figure 5.30, and the watermarked images with PSNR fixed to $28dB$ to the original lena are shown in Figures to 5.37. The original and watermarked images of Cameraman, Barbara and Peppers are not showed in this thesis, only for the purpose of saving pages. The experimental conditions are similar to what we used for Corel images: Two bits are embedded into a 8×8 block and 62 DCT coefficients are used in each block.

Figure 5.30: Original Image

Figure 5.31: Watermarked lena by traditional DM

Figure 5.32: Watermarked lena by Oostveen

Figure 5.33: Watermarked lena by DM-MW

Figure 5.34: Watermarked lena by RDM-62-L2-Norm

Figure 5.35: Watermarked lena by RDM

Figure 5.36: Watermarked lena by STDM

Figure 5.37: Watermarked lena by STDM-OptiMW-SS

5.6.1 Subjective Quality Assessment

In this thesis, we have measured the quality of the watermarked image by both subjective measurements (such as DWR and PSNR) and Watson's distance defined by the perceptual model. In addition to these, we performed another quality assessment by a user trial. This experiment is done as follows:

- Send the original 512×512 256 greyscale Lena image showed in 5.30 (but with the raw BMP format) to a user
- Mark the watermarked images using different methods, showed in Figures 5.32 to 5.37, as Image 1, 2, ..., 7. Send all these images to users and the users have no knowledge of which picture is generated by which method.
- Have users score the watermarked image in the way that: 10 points means the watermarked image has no visual differences from the original image, 9 points means it has trivial perceptual differences, 8 points means it gets worse. The

lower the score it has, the bigger the difference is.

- Collect the forms with scores that users have filled in.

We have sent out trial requests to 105 users and got 93 valid forms back. The average scores for all valid results are showed in Table .

Scheme	Figure index	Average User subjective score
DM	5.31	6.4
Oostveen	5.32	5.6
DM-MW	5.33	8.7
RDM-62-L2-Norm	5.34	6.3
RDM	5.35	3.7
STDM	5.36	6.5
STDM-OptiMW-SS	5.37	8.2

Table 5.5: User trial

Results in Table shows Figures 5.37 and 5.33 have higher score of 8.2 and 8.7. In contrast, the watermarked image produced by traditional DM has a score of 6.4, the previously proposed method of RDM and Oostveen achieved a score of 3.7 and 5.6, respectively. It indicates that users consider Figures 5.37 and 5.33 generated by our methods to be visually closer to the original image, compared with other images. This confirms that watermarked images based on a perceptual model are visually more pleasant than previous QIM methods. Please note that this is achieved by fixing the strength of watermarking in terms of PSNR.

5.6.2 Experimental results against amplitude scaling and JPEG

For the standard images (Lena, Cameraman, Barbara and Peppers) watermarked using different methods, we firstly test their performances against amplitude scaling. The results are shown in Figure 5.38.

Once again, in Figure 5.38, our methods based on the modified Watson's model, such as DM-MW, RDM-MW, STDM-OptiMW-SS have lower BER than previous methods.

We then give the experimental results of the watermarked images against JPEG compression. The BERs are shown in Figure 5.39.

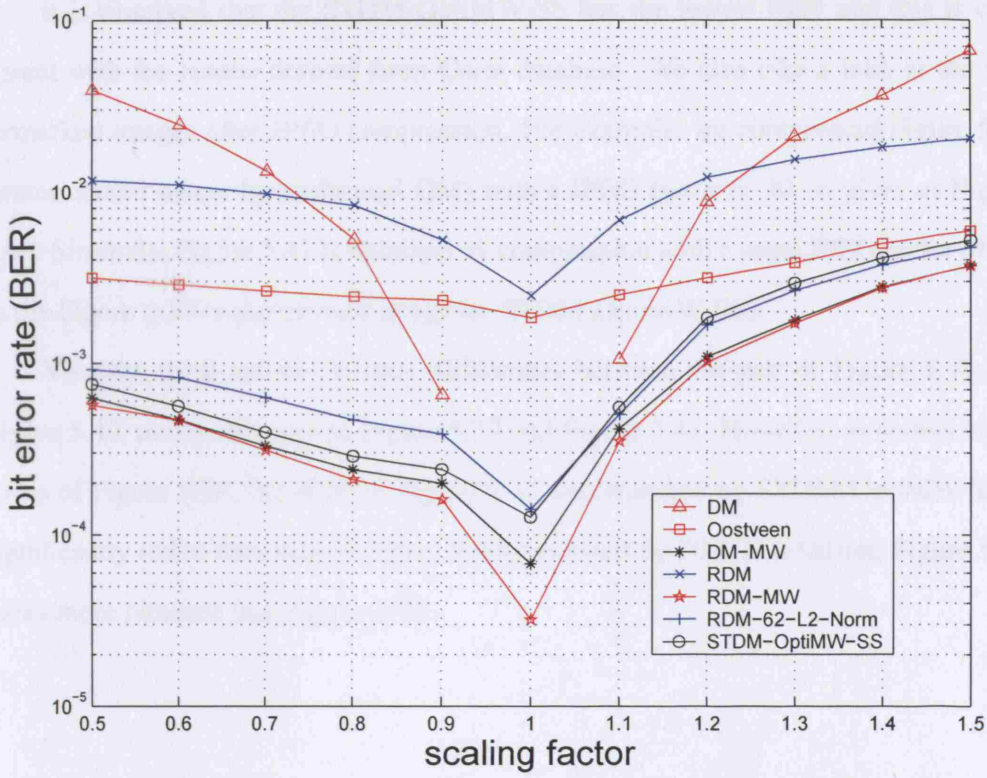


Figure 5.38: BER versus amplitude scaling using standard images

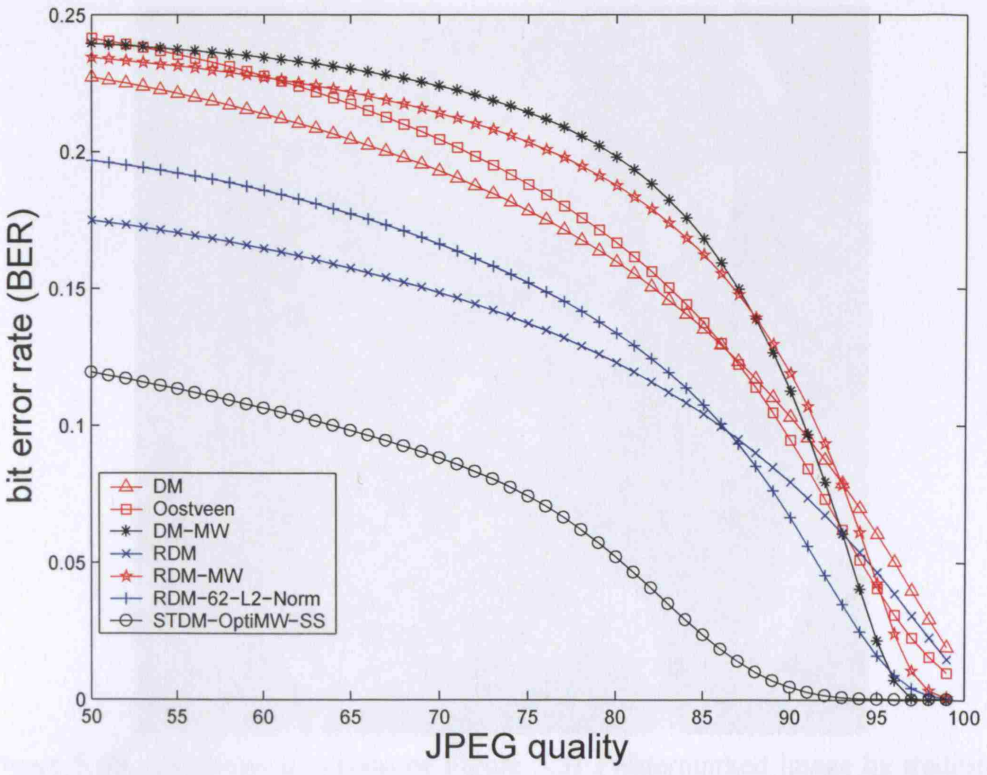


Figure 5.39: BER as a function of JPEG quality using standard images

It is observed that the STDM-OptiMW-SS has the lowest BER and this is consistent with the results derived from Corel database. We also take a look at the watermarked images after JPEG compression. For example, we compressed Figure 5.31 (watermarked image by traditional DM) with a JPEG factor of 80. It gives us Figure 5.40. Similarly, Figure 5.41 is obtained by compression with a same JPEG factor of 80, to the Figure 5.37 (watermarked image by STDM-OptiMW-SS).

Visually, there are no obvious differences between the pair of Figure 5.31 and Figure 5.40, neither the pair of Figure 5.37 and Figure 5.41. However, as shown in the curve of Figure 5.39, the BER of Figure 5.41 watermarked by STDM-OptiMW-SS is significantly lower than that of Figure 5.40 generated by DM. In addition, Figure 5.41 looks more pleasant than Figure 5.40.

Figure 5.40: Compressed version of Figure 5.31 (Watermarked image by traditional DM),with JPEG factor of 80

Figure 5.41: Compressed version of Figure 5.37 (Watermarked image by STDM-OptiMW-SS),with JPEG factor of 80

Chapter 6

Conclusion

Quantization index modulation (QIM), originally proposed by Chen and Wornell, is a form of digital watermarking based on the framework of communications with side information. Its resulting lattice codes exhibit high capacity and computational simplicity. As a result, QIM has received significant attention within the watermarking community. However, this approach has some limitations. For example, it introduces distortion which is not adaptive to the local features of the host signal and thus may be perceptually noticeable. More significant limitations of QIM algorithm are their extreme sensitivity to valumetric scaling, e.g., changes of amplitude, and re-quantization, e.g., lossy compression. This thesis examined a variety of ways in which these limitations of QIM could be overcome. Our contributions are summarised in Section 6.1. In addition, some directions of future work are given in 6.2.

6.1 Summary of Contributions

We have proposed several modifications to dither modulated QIM. The first method, DM-W, uses Watson's perceptual model to adaptively change the quantization step size for dither modulation in order to improve fidelity. Experimental results confirmed that for the same document-to-watermark ratio (DWR), the Watson distance is reduced by over 80%. This improvement is achieved while simultaneously improving the robustness in high noise regimes.

Next, we modified Watson's perceptual model so that the adaptive QIM scheme

(DM-MW) is invariant to valumetric scaling. Experimental results demonstrated that when using soft decision decoding and an embedding rate of 1-bit-per-32 pixels at a DWR = 35dB, the BER does not exceed 1% over a scale range of 0.5 to 1.0. When the scaling factor increases to 1.5, the BER is still less than 7%. While there is a small degradation in fidelity compared with DM-W, the perceptual distortion introduced by DM-MW is only about one quarter of the regular DM (as measured by Watson distance). The distortion is also much lower than the adaptive method of Oostveen *et al.* [OKS04], and standard rational dither modulation (RDM).

This algorithm, DM-MW, implicitly assumes that the slacks calculated at the embedder and detector are the same, despite the modifications to the DCT coefficients introduced by the embedder. This is a source of error which is eliminated by using rational dither modulation with a modified Watson measure (RDM-MW). The adaptive step size for the current block is now based on perceptual estimates from the previously watermarked, neighboring block. This guarantees that the slacks calculated at the embedder and detector are identical (prior to any distortions between embedding and detecting). Of course, using the perceptual slacks calculated from the previously watermarked neighborhood to affect the step size in the current block, is likely to introduce some distortions and a further small degradation in fidelity, compared with both DM-W and DM-MW, is observed. The Watson distance is slightly increased from 9.4 of DM-MW to 10.2 of RDM-MW. However, this degradation is small and the use of RDM-MW reduces the bit error rate from 0.08% of DM-MW to 0.03% when the embedding efficiency for 1,000 images is examined.

Experimental results also demonstrated that the performance of DM-W, DM-MW and RDM-MW degrade more smoothly with the addition of white Gaussian noise, compared with traditional dither modulation (DM). This provides superior performance in high noise regimes.

We also tested the performance of these algorithms to addition/subtraction of a constant luminance value. Experimental results comparing the performance with that of Oostveen *et al.* [OKS04] indicate the all three algorithms are significantly more

robust. However, standard DM in the DCT domain (DM) is the most robust.

Our experimental investigations also revealed that a significant source of residual errors is due to rounding which occurs after computing the inverse DCT and converting to integer pixel values. These rounding errors may cause detection errors when the step size used for QIM is small. Rounding errors are, of course, a form of quantization, and most QIM, DM and DC-QIM methods are sensitive to re-quantization which can also occur as a result of analog-to-digital conversion and JPEG compression. Sensitivity to JPEG compression was experimentally examined. We observed that all the algorithms considered above, such as DM, DM-W, DM-MW and RDM-MW, are very sensitive to JPEG compression. If we maintain a fixed DWR, we generally observe that the perceptually-based algorithms perform worse than those that do not use a perceptual model. However, if we fix the perceptual distortion rather than DWR, then the perceptually-based algorithms provide significant improvements in performance.

Quantization index modulation methods are fragile to valumetric scaling and re-quantization. While considerable work has been directed to improving the robustness to scaling, the issue of re-quantization has received much less attention. This is surprising since re-quantization commonly occurs due to lossy compression, numerical rounding and analog-to-digital conversion.

Of the number of QIM variants, spread transform dither modulation exhibits most robustness to re-quantization. We have described how a perceptual model can be incorporated into the STDM framework and proposed several extensions to STDM. In the first method, referred to as STDM-W (STDM Watson), the random projection vector is replaced with the vector representing the magnitudes of the Watson's slacks associated with each DCT coefficient. This has the effect of directing the change to the host signal to those areas of the signal where the changes will be least noticeable. Since the projection vector is now image dependent, the detector must also be able to estimate the slack for each image. Although the very act of watermark embedding may slightly change these slack values, experimental results indicate that these changes are sufficiently small to permit very good detector performance. Meanwhile, STDM-

W considerably reduces the perceptual distortion from 41 to 8.7, compared with the STDM.

By using a modified Watson's perceptual model which scale linearly with valumetric scaling and using this not only to select the projection vector but also to determine the quantization step size, the new method STDM-MW-SS is able to significantly improve both the fidelity and robustness to valumetric scaling, compared with original STDM. Experimental results on 1000 images confirmed these designs.

However, experiments revealed an unexpected sensitivity to JPEG compression. This was due to the modified perceptual model over-estimating the slack values for the high frequency DCT coefficients. As a results, more of the watermark energy was placed in these regions, which are very sensitive to JEPG compression. This problem was resolved by introducing a piecewise linear model that attenuated the slack estimates. The resulting algorithm, STDM-OptiMW-SS exhibits very good fidelity, and is very robust to both JPEG compression and valumetric scaling. At a fixed Watson Distance of 39, and using an embedding rate of $1/32$, STDM-OptiMW-SS provide very low BER (less than 0.5%) when images amplitude are scaled down. The BER never exceeds 7% when images amplitude are scaled up with a factor in the range of 1.0 to 1.5. Under the same experimental conditions, the robustness of STDM-OptiMW-SS to JPEG compression was also examined. The BER does not exceed 20% until the JPEG quality factor is less than 64%, and the BER is still lower than 24% even for JPEG quality factor of 50%.

We believe these results are important as it is imperative that watermarks be robust (preferably invariant) to valumetric scaling and re-quantization if they are to be applied in copy protection applications where lossy compression, D-to-A and A-to-D conversion and changes in brightness/volume are common.

6.2 Future Work

This section discusses some potential directions to extend the results of this thesis.

- In the algorithm of STDM-OptiMW-SS, the step size calculated in an embed-

der and a detector is not identical because embedding alters the host signal, and consequently, perceptual slacks. Further improvements may be possible if this source of error can be reduced.

- The soft-decision decoding used in our methods equally weight and accumulate the distance obtained from every signal sample for detection. The robustness to JPEG compression may be further improved if the different DCT coefficients are weighted appropriately.

- The algorithms we have described are block based. As such, they are robust to spatially varying valumetric changes, if such changes occur on a block-by-block basis. However, highly non-linear valumetric changes, such as gamma correction, remain problematic. This is an area of future work.

- This thesis mainly discussed results based on 256 greyscale images, the extension of the proposed algorithms to colour images is also a future direction.

Appendix A

Author's Publications

Jounral paper

- Qiao Li, Ingemar J. Cox, "Using perceptual models to improve fidelity and provide resistance to valumetric scaling for Quantization Index Modulation Watermarking", IEEE Transaction on Information Forensics and Security, VOL.2, NO.2, June 2007

Conference papers

- Qiao Li, Ingemar J. Cox, "Improved spread transform dither modulation using a perceptual model: robustness to amplitude scaling and jpeg compression", IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), April, 2007, Honolulu, U.S.A.

- Qiao Li, Gwenael Doerr, Ingemar J. Cox, "Spread transform dither modulation using a perceptual model", IEEE International Workshop on Multimedia Signal Processing (MMSP), Oct., 2006, Victoria, Canada

- Qiao Li, Ingemar J. Cox, "Rational dither modulation watermarking using a perceptual model", IEEE International Workshop on Multimedia Signal Processing (MMSP), Nov., 2005, Shanghai, China.

- Qiao Li, Ingemar J. Cox, "Using perceptual models to improve fidelity and provide invariance to valumetric scaling for quantization index modulation wa-

termarking", IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), March, 2005, Philadelphia, U.S.A.

Appendix B

Glossary

BER Bit error rate, the percentage of bits that have errors relative to the total number of bits embedded as a watermark message, defined in Section 2.3.

DWR Document-to-watermark ratio, defined in Equation (2.2).

JPEG QF JPEG *quality factor*, described in Section 3.5.2.

MSE Mean square error, defined in Equation (2.1).

Watson distance A measurement for the fidelity of a distorted image to its original image, described in Equation (2.24).

Bibliography

[APGM05] A. Abrardo, M. Barni F. Perez-Gonzalez, and C. Mosquera. Trellis-coded rational dither modulation for digital watermarking. In *in Proceedings of the 4th International Workshop on Digital Watermarking*, volume LNCS 3710, pages 351–360, September 2005.

[Bar] David Barry. Aggies help combat digital piracy through watermarking technology. HTML.

[Bas05] Patrick Bas. A quantization watermarking technique robust to linear and non-linear valumetric distortions using a fractal set of floating quantizers. In *7th Information Hiding Workshop*, Barcelona, Spain, June 2005.

[BCK+99] J. A. Bloom, I. J. Cox, T. Kalker, J-P Linnartz, M. L. Miller, and B. Traw. Copy protection for DVD video. *Proc. IEEE*, 87(7):1267–1276, 1999.

[Bet] Watermark helps track movie uploader. HTML.

[BG99a] B.Chen and G.Wornell. Dither modulation: a new approach to digital watermarking and information embedding. In *Proc. of SPIE: Security, Steganography, and Watermarking of Multimedia Contents* , volume 5306, pages 342–353, San Jose, California, USA, January 1999.

[BG99b] B.Chen and G.Wornell. An information–theoretic approach to the design of robust digital watermarking systems. In *Int. Conf. on Acoustics,Speech and Signal Processing*, Phoenix, USA, March 1999.

[BG01a] B.Chen and G.W.Wornell. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Informatin Theory*, 47:1423–1443, May 2001.

[BG01b] B.Chen and G.W.Wornell. Quantization Index Modulation methods for digital watermarking and information embedding of multimedia. *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology, Special Issue on Multimedia Signal Processing*, pages 7–33, Feb 2001.

[BGML96] W. Bender, D. Gruhl, N. Morimoto, and A. Lu. Techniques for data hiding. *IBM Systems Journal*, 35(3/4):313–336, 1996.

[BS99] Z. Bahrav and D. Shaked. Watermarking of dithered halftoned images. In *Proc. SPIE Conf. on Security and Watermarking of Multimedia Data*, volume 3657, pages 307–316, 1999.

[CKLS96] I.J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for images, audio and video. In *IEEE Int. Conference on Image Processing*, volume 3, pages 243–246, 1996.

[CKLS97] I.J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Trans. on Image Processing*, 6(12):1673–1687, 1997.

[CM97] I.J. Cox and M. L. Miller. A review of watermarking and the importance of perceptual modeling. In *Proceedings of SPIE, Human Vision & Electronic Imaging II*, volume 3016, pages 92–99, 1997.

[CMB01] I. J. Cox, M. L. Miller, and J. A. Bloom. *Digital Watermarking*. Morgan Kaufmann, 2001.

[CMM99] I. J. Cox, M. L. Miller, and A. McKellips. Watermarking as communications with side information. *Proc. IEEE*, 87(7):1127–1141, 1999.

[CMT+99] D. Coppersmith, F. Mintzer, C. Tresser, C. W. Wu, and M. M. Yeung. Fragile imperceptible digital watermark with privacy control. In *Security and Watermarking of Multimedia Contents*, volume SPIE-3657, pages 79–84, 1999.

[Cos83] M. Costa. Writing on dirty paper. *IEEE Trans. Inform. Theory*, 29(3):439–441, May 1983.

[CPR99] Jim Chou, S. Sandeep Pradhan, and Kannan Ramchandran. On the duality between distributed source coding and data hiding. *Thirty-third Asilomar conference on signals, systems, and computers*, 2:1503–1507, 1999.

[CW00] B. Chen and G. W. Wornell. Preprocessed and postprocessed quantization index modulation methods for digital watermarking. In *Security and Watermarking of Multimedia Contents II*, volume SPIE-3971, pages 48–59, 2000.

[Dec01] Steve Decker. Engineering considerations in commercial watermarking. *IEEE Communications Magazine*, To be published, 2001.

[EBG02] J.J. Eggers, R. Bauml, and B. Girod. Estimation of amplitude modifications before SCS watermark detection. In *Proc. SPIE: Security, Steganography, and Watermarking of Multimedia Contents*, pages 387–398, San Jose, California, USA, January 2002.

[EBTG03] J.J. Eggers, R. Bauml, R. Tzschoppe, and B Girod. Scalar Costa Scheme for information embedding. *Signal Processing, IEEE Transactions on*, 51(4):1003–1019, April 2003.

[EKK99] F. Ergun, J. Kilian, and R. Kumar. A note on the limits of collusion-resistant watermarks. In *Advances in Cryptology – EUROCRYPT '99*, pages 140–149. Springer-Verlag, 1999.

[FA00] M. S. Fu and O. C. Au. Data hiding for halftone images. In *Proc. SPIE Conf. on Security and Watermarking of Multimedia Data*, volume 3971, pages 228–236, 2000.

[FKK04] C. Fei, D. Kundur, and R. Kwong. Analysis and design of watermarking algorithms for improved resistance to compression. *IEEE Transactions on Image Processing*, 13(2):126–144, February 2004.

[Gir93] Bernd Girod. What's wrong with mean-squared error? In A. B. Watson, editor, *Digital Images and Human Vision*, chapter 15, pages 207–220. MIT Press, 1993.

[HH05a] M. Hagmuller and H. Hering. Speech watermarking for air traffic control. In *Eurocontrol Experimental Centre Note no. 05/05*, 2005.

[HH05b] K Hofbauer and H. Hering. Digital signatures for the analogue radio. In *Proceedings of the 5th NASA Integrated Communications Navigation and Surveillance (ICNS) Conference and Workshop*, 2005.

[HHK03] H. Hering, M. Hagmuller, and G. Kubin. Safety and security increase for air traffic management through unnoticeable watermark aircraft identification tag transmitted with the vhf voice communication. In *Digital Avionics Systems Conference DASC'03, The 22nd*, volume 1, pages 4.E.2 – 41–10, Oct. 2003.

[HHK04] H. Hering, M. Hagmuller, and G. Kubin. Speech watermarking for air traffic control. In *12th European Signal Processing Conference, (EUSIPCO)*, pages 1653–1656, 2004.

[HK99] Frank Hartung and Martin Kutter. Multimedia watermarking techniques. *Proc. IEEE*, 87(7):1079–1107, 1999.

[HZG04] Anthony T. S. Ho, Xunzhan Zhu, and Yong Liang Guan. Image content authentication using pinned sine transform. *EURASIP Journal on Applied Signal Processing*, (14):2174–2184, 2004.

[KLM+97] J. Kilian, F. T. Leighton, L. R. Matheson, T. Shamoon, and R. E. Tarjan. Resistance of watermarked documents to collusional attacks. Technical Report TR 97-167, NEC Research Institute, 1997.

[LC00] C-Y Lin and S-F Chang. Semi-fragile watermarking for authenticating JPEG visual content. In *Proc of SPIE, Security and Watermarking in Multimedia Contents II*, volume 3971, 2000.

[LC05a] Qiao Li and Ingemar J. Cox. Rational dither modulation watermarking using a perceptual model. In *IEEE International Workshop on Multimedia Signal Processing(MMSP)*, Shanghai, China, Oct. 2005.

[LC05b] Qiao Li and Ingemar J. Cox. Using perceptual models to improve fidelity and provide invariance to valumetric scaling for quantization index modulation watermarking. In *IEEE Int. Conf. on Acoustics,Speech and Signal Processing(ICASSP)*, Philadelphia, USA, March 2005.

[LKKM03] Kiryung Lee, Dong Sik Kim, Taejeong Kim, and Kyung Ae Moon. Em estimation of scale factor for quantization-based audio watermarking. In *Digital Watermarking,Second International Workshop,IWDW 2003*, Seoul, Korea, Oct. 2003.

[LPD00] E. T. Lin, C. I. Podilchuk, and E. J. Delp. Detection of image alterations using semi-fragile watermarks. In *SPIE*, 2000.

[LS04] R. L. Lagendijk and Ivo D. Shterev. Estimation of attacker's scale and noise variance for qim-dc watermark embedding. In *IEEE Int. Conf. on Image Processing*, Singapore, Oct. 2004.

[LT98] J. P. M. G. Linnartz and J.C. Talstra. MPEG PTY-marks: Cheap detection of embedded copyright data in DVD-video. In *Proceeding of ESORICS98 5th European Symposium on Research In Computer Security*, 1998.

[MA42] Hedy Kiesler Markey and George Antheil. Secret communication system. Technical Report 2,292,387, United States Patent, 1942.

[MCB00] M. L. Miller, I. J. Cox, and J. A. Bloom. Informed embedding: Exploiting image and detector information during watermark insertion. In *IEEE International Conference on Image Processing*, September 2000.

[MDC02] M. L. Miller, G. J. Doerr, and I. J. Cox. Dirty-paper trellis codes for watermarking. In *IEEE Int. Conf. on Image Processing*, Rochester,USA, Sep. 2002.

[MDC04] M. L. Miller, G. J. Doerr, and I. J. Cox. Applying informed coding and embedding to design a robust high-capacity watermark. *IEEE Transactions on Image Processing*, 13:792–807, 2004.

[MF03] H.S. Malvar and D.A.F. Florencio. Improved spread spectrum: a new modulation technique for robust watermarking. *Signal Processing, IEEE Transactions on*, 51(4):898– 905, April 2003.

[NP84] N.S.Jayant and P.Noll. *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Prentice-Hall, 1984.

[OKS04] J.C. Oostveen, A.A.C. Kalker, and M. Staring. Adaptive quantization watermarking. In *Proc. of SPIE:Security, Steganography, and Watermarking of Multimedia Contents VI*, volume 5306, pages 296–303, San Jose, California, USA, January 2004.

[Ó RuanaidhDB96] J. J. K. Ó Ruanaidh, W.J. Dowling, and E.M. Boland. Phase watermarking of digital images. *Proc. IEEE Int. Conf. on Image Processing*, III:239–242, 1996.

[Ó RuanaidhP98] J. J. K. Ó Ruanaidh and T. Pun. Rotation, scale and translation invariant spread spectrum digital image watermarking. *Signal Processing*, 66(3):303–317, 1998.

[PAK99a] F.A. Petitcolas, R. J. Anderson, and M.G. Kuhn. Information hiding: a survey. *Proceedings of the IEEE, Special Issue on Identication and Protection of Multimedia Information*, 87(7):1062–1078, July 1999.

[PAK99b] Fabien A. P. Petitcolas, Ross Anderson, and Markus G. Kuhn. Information hiding - a survey. *Proc. IEEE*, 87(7):1062–1077, 1999.

[PGBAM04] F. Perez-Gonzalez, M. Barni, A. Abrardo, and C. Mosquera. Rational dither modulation: a novel data-hiding method robust to value-metric scaling attacks. In *IEEE International Workshop on Multimedia Signal Processing*, Siena, Italy, Sep. 2004.

[PGBH03] F. Perez-Gonzalez, F. Balado, and J. R. Hernandez. Performance analysis of existing and new methods for data hiding with known-host information in additive channels. *Signal Processing, IEEE Transactions on*, 51(4):960–980, April 2003.

[PGCB03] F. Pérez-Gonzàlez, P. Comesaña, and F. Balado. Dither-modulation data hiding with distortion-compensation: Exact performance analysis and an improved detector for JPEG attacks. In

Proceedings of the IEEE International Conference on Image Processing, volume II, pages 503–506, September 2003.

[PGMBA05] F. Pérez-Gonzàlez, F. Mosquera, M. Barni, and A. Abrardo. Rational dither modulation: A high-rate data-hiding method invariant to gain attacks. *IEEE Transactions on Signal Processing, Supplement on Secure Media*, 53(10):3960–3975, October 2005.

[PSM82] R. L. Pickholtz, D. L. Schilling, and L. B. Milstein. Theory of spread-spectrum communications - a tutorial. *IEEE Trans. on Communications*, 30(5):855–884, 1982.

[PZ98] Christine I. Podilchuk and Wenjun Zeng. Image-adaptive watermarking using visual models. *IEEE Journal of Selected Areas in Communication*, 16(4):525–539, 1998.

[RM96] R.Zamir and M.Feder. On lattice quantization noise. *IEEE Transactions on Information Theory*, pages 1152–1159, July 1996.

[SC05] Qibin Sun and Shih-Fu Chang. A secure and robust digital signature scheme for jpeg2000 image authentication. *Multimedia, IEEE Transactions on*, 7(3):480–494, June 2005.

[Sch64] Leonard Schuchman. Dither signals and their effect on quantization noise. *IEEE Trans. Comm. Tech.*, 12:162–165, 1964.

[SF96] D.A. Silverstein and J.E. Farrell. The relationship between image fidelity and image quality. In *IEEE Int. Conf. on Image Processing*, pages 881–884, Lausanne, Switzerland, Sep. 1996.

[SL06] I.D. Shterev and R. Lagendijk. Amplitude scale estimation for quantization-based watermarking. *IEEE Transactions on Signal Processing*, 54:4146– 4155, Nov. 2006.

[SLH04] Ivo D. Shterev, Inald L. Lagendijk, and Richard Heusdens. Statistical amplitude scale estimation for quantization-based watermarking. In *Proc. of SPIE:Security, Steganography, and Watermarking of Multimedia Contents VI*, volume 5306, pages 796–804, San Jose, California, USA, Jan. 2004.

[Sto96] H. S. Stone. Analysis of attacks on image watermarks with randomized coefficients. Technical Report TR 96-045, NEC Research Institute, 1996.

[SZT96] M. D. Swanson, B. Zhu, and A. H. Tewfik. Transparent robust image watermarking. In *Proc. IEEE Int. Conf. on Image Processing*, volume 3, pages 211 –214, 1996.

[the] Using digital watermarking to fight against content piracy. HTML.

[TWWL03] W. Trappe, Min Wu, Z.J. Wang, and K.J.R Liu. Anti-collusion fingerprinting for multimedia. *Signal Processing, IEEE Transactions on*, 51(4):1069– 1087, April 2003.

[VHBP99] S. Voloshynovskiy, A. Herrigel, N. Baumgaetner, and T. Pun. A stochastic approach to content adaptive digital image watermarking. In *Third International Workshop on Information Hiding*, 1999.

[Vit95] A. J. Viterbi. *CDMA: principles of spread spectrum communications*. Addison Wesley Longman Inc., 1995.

[Wat93a] Andrew B. Watson. DCT quantization matrices optimized for individual images. *Human Vision, Visual Processing, and Digital Display IV*, SPIE-1993:202–216, 1993.

[Wat93b] Andrew B. Watson. Visually optimal DCT quantization matrices for individual images. *Proc. IEEE Data Comp. Conf.*, pages 178–187, 1993.

[WB06] Zhou Wang and Alan C. Bovik. *Modern Image Quality Assessment.* Morgan & Claypool Publishers, 2006.

[WBSS04] Z Wang, AC Bovik, HR Sheikh, and EP Simoncelli. image quality assessment: From error visibility to structural similarity. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 13(4):600–612, Apr. 2004.

[WD96] Raymond B. Wolfgang and Edward J. Delp. A watermark for digital images. In *Proceedings of the 1996 International Conference on Image Processing*, volume 3, pages 219–222, 1996.

[WD99] Raymond B. Wolfgang and Edward J. Delp. Fragile watermarking using the VW2D watermark. In *Security and Watermarking of Multimedia Contents*, volume SPIE-3657, pages 204–213, 1999.

[WK00] S.-G. Wang and K. T. Knox. Embedding digital watermarks in halftone screens. In *Proc. SPIE Conf. on Security and Watermarking of Multimedia Data*, volume 3971, pages 218–227, 2000.

[WL98] Min Wu and Bede Liu. Watermarking for image authentication. In *IEEE Int. Conf. on Image Processing*, volume 2, pages 437 –441, 1998.

[WMC04] C. K. Wang, M. L. Miller, and I. J. Cox. Using perceptual distance to improve the selection of dirty paper trellis codes for watermarking. In *IEEE Int. Workshop on Multimedia Signal Processing*, Siena, Italy, Sep. 2004.

[WPD99] Raymond B. Wolfgang, C. I. Podilchuk, and E. J. Delp. Perceptual watermarks for digital images and video. *Proc. of the IEEE*, 87(7):1108–1126, 1999.